MESG

MESTRADO EM ENGENHARIA DE SERVIÇOS E GESTÃO

Developing visualisations for urban mobility data: a user-centred design approach

Thiago Sobral Marques da Silva

Dissertação de Mestrado

Orientador na FEUP: Prof. Maria Teresa Galvão Dias

Universidade do Porto

Faculdade de Engenharia

FEUP

Faculdade de Engenharia da Universidade do Porto

2014-07-04

IBM Centers for Advanced Studies

Enterprise Engineering and Management

To my parents.

Resumo

Desde as últimas décadas, observa-se um aumento notório dos níveis de mobilidade em áreas urbanas, causados por fatores como evoluções econômicas e mudanças políticas. Esse fenómeno evidencia a necessidade de se identificar tendências de mobilidade urbana periodicamente. As entidades que se preocupam em analisar tais tendências podem contribuir positivamente para a qualidade de mobilidade urbana, obter vantagens competitivas e descobrir oportunidades de inovação.

Os Sistemas Inteligentes de Transportes configuram-se como uma forma de melhorar o transporte através da integração entre a infraestrutura de transporte e as tecnologias de comunicação e informação. Esses sistemas geram grandes quantidades de dados que geralmente não são explorados em profundidade pelos intervenientes que os detêm. Da mesma forma, a crescente evolução da área de Visualização da Informação tem oferecido novas técnicas que auxiliam a extração de conhecimento a partir de conjuntos de dados de larga escala. As visualizações podem oferecer contribuições poderosas para autoridades e empresas ligadas ao setor de transporte, permitindo melhor compreensão acerca dos padrões e tendências de mobilidade de uma cidade. Embora a Visualização da Informação seja reconhecida como uma área centrada no utilizador, ainda é possível observar a escassez de abordagens efetivas centradas no utilizador para o desenvolvimento de visualizações. O envolvimento de utilizadores potenciais durante todas as fases de desenho de um sistema é crucial, dado que estes serão os maiores beneficiados durante as suas rotinas e decisões diárias.

A fim de reduzir esta lacuna e enfatizar a necessidade de se explorar o grande volume de dados de mobilidade urbana existente, este estudo empírico propõe uma abordagem completamente centrada no utilizador para o desenvolvimento de visualizações para o setor de mobilidade urbana, identificando que tipos de visualizações podem contribuir efetivamente para extração do conhecimento, dependendo do tipo de dados em causa. Utilizou-se diferentes tipos de dados de mobilidade urbana da cidade do Porto, os quais foram obtidos através de diversas fontes como contadoras de tráfego, bilhética e uma aplicação móvel que fornece informações em tempo real sobre transportes públicos. A metodologia compreendeu uma caracterização formal dos dados obtidos e a modelação temporal dos mesmos. Posteriormente, iniciou-se o processo de desenvolvimento de visualizações altamente iterativo, focado na prototipagem rápida e avaliação através de testes de usabilidade exploratórios com peritos em mobilidade urbana.

Finalmente, os peritos constataram o potencial dos conjuntos de dados que possuem e propuseram melhorias valiosas para as técnicas de visualização que foram aceites. Pretende-se que os resultados deste estudo evidenciem a importância de se envolver continuamente os utilizadores durante o processo de desenvolvimento de visualizações para o setor de mobilidade urbana e outros possíveis domínios onde também sejam aplicáveis.

Abstract

Mobility levels have increased notoriously in urban areas on the last decades driven by economic and political changes. This change poses the need of identifying mobility trends periodically. Entities that are concerned with those trends can positively contribute to citizens' perception of urban mobility quality, gain competitive advantage and uncover innovation opportunities.

Intelligent Transport Systems have been one of the approaches for improving transport by integrating infrastructure with information and communication technologies. They generate massive amounts of raw data that usually are not explored in-depth by the ones who own them. Likewise, the evolution of Information Visualisation provides new techniques for assisting knowledge extraction from large datasets. Visualisations can be a powerful asset for authorities and transport related services for identifying mobility trends and gaining in-depth understanding about a city's dynamics. Although Information Visualisation is acknowledged as user-centric, current research is still scarce in applying thorough user-centred design approaches. Involving potential users during all phases of any design process is crucial, as they are the ones who will benefit from visualisations on their daily routines and decisions.

In order to reduce this gap and emphasise the need of exploring data, this empirical research propose a thorough user-centred approach for developing visualisations for the urban mobility domain, in order to assess what types of visualisations can effectively support knowledge extraction depending on the type of data that is concerned. Different types of urban mobility data from Porto were used, which were extracted from different sources including traffic counter sensors, ticketing and a mobile application that provides real time information about public transport. The methodology comprised a formal characterisation of data and time modelling. After, a highly iterative design process started, which focused on rapid prototyping and evaluation through exploratory usability tests with experts in urban mobility.

Finally, experts realised the hidden potential of datasets and proposed valuable improvements for the approved visualisation techniques. The results of this research are expected to enlighten researchers and practitioners about the need of actively involving users, as well as to serve as a reference for developing visualisations for the urban mobility domain.

Acknowledgements

I am deeply grateful to Professors Teresa Galvão and José Luís Borges for their confidence, motivation and continuous support. This work immensely benefited from our fruitful conversations about urban mobility and Information Visualisation. I feel honoured for the opportunity of doing this research at IBM Centre for Advanced Studies – Portugal.

I extend my gratitude to Professors João Falcão e Cunha and Lia Patrício, directors of the Master in Services Engineering and Management, the academic and secretariat staff for their friendliness and dedication to this program.

This work benefits from the kind support and participation of the staff of OPT S.A, the Mobility Division of the Porto City Hall, STCP S.A. and other urban mobility experts from Porto who provided datasets and participated in this research. Thank you very much for collaborating with the success of this research.

Words cannot describe how grateful I am for my family's love and support during these two years distant from home; especially for my parents who always supported my academic career and the decision to study abroad, always pouring cheering words in times when the fact of being far from home were harder to endure.

Last but obviously not least, I express my deep gratitude for the precious friendships I settled here in Portugal and the good memories I will carry from now on. I am grateful for their care and continuous support. Thank you for making me feel at home in Porto.

Index of Contents

Index of Figures

Index of Tables

Chapter 1 Introduction

Mobility levels have increased notoriously in urban areas on the last decades driven by economic and political changes. It is forecast that 6.3 billion people will live in urban areas by 2050, which sum up to 70% of the world population (United Nations 2012). Moreover, factors such as lifestyle, longer life expectancy and migration have direct impact on urban mobility behaviour and urban infrastructure demands, including public transport. Hence, authorities and transport services should identify mobility trends periodically, analyse and (re)shape their operations based on those changes. These attitudes can contribute to citizens' perception of service quality, offer competitive advantage to transport services and uncover opportunities for innovation.

Intelligent Transport Systems (ITS) have been one of the approaches for improving transport by integrating Information and Communication Technologies (ICT) into transport infrastructure. ITS allows retrieving data such as travel behaviour, ticketing and traffic flows, which is made possible by telematics, electronic ticketing, travel information systems and traffic sensors. Through the lens of Big Data, those data sources are powerful for understanding the dynamics of urban mobility within cities.

Likewise, the evolution of Information Visualisation brought new forms of extracting knowledge from large datasets, and has been applied to diverse areas. The variety of techniques and tools allows the continuous development of creative and scalable ways of visualizing data. Nonetheless, the richness of technological features does not waive understanding the domain of application and the users who will take benefit from visualisations.

Notwithstanding that Information Visualisation is considered to be 'user-centred', current research is still scarce in providing examples of thorough approaches that involve users actively during the design process. Most studies found on literature focus on presenting and demonstrating visualisation techniques, either not emphasising evaluation with potential users or not evaluating at all. This puts to question how adequate those techniques actually are, considering the perspective of those who will use them. Users cannot extract knowledge from flawed or context-inadequate visualisations. Also, authorities and transport services typically do not explore the data they own at its full extent, missing a potential opportunity to understand the dynamics of a city.

1.1 Objective and research questions

The aforementioned scenario motivates this research, which encompasses the areas of Information Visualisation and User-Centred Design (UCD) within the Urban Mobility domain. It adopts a thorough user-centred approach for developing visualisation techniques oriented to urban mobility experts who belong to strategic and operational levels of transport related services and authorities. We took advantage of available datasets from Porto's bus ticketing system; travel planning (retrieved from Move-me[1]) and traffic counter sensors.

This research addresses the following research objective and its associated research questions:

O1. Design and evaluate the adequacy of visualisation techniques oriented to the urban mobility domain by following a user-centred design approach.

RQ1. Can the proposed user-centred design methodology produce meaningful visualisations for the urban mobility domain?

RQ2. Depending on the data that is concerned, which visualisation techniques support knowledge extraction and collaborate on decision-making routines of urban mobility experts?

We adopted a user-centred methodology that emphasises rapid prototyping and continuous evaluation, involving experts in urban mobility from public transport services, city authorities and researchers in order to evaluate prototypes, retrieve feedback and elicit user requirements. The involvement of experts ensures that the visualisations are consistent and reliable regarding the activities they will support. Also, given that experts are the ones who will make extensive use of visualisations, it is appropriate to prototype visualisations according to their needs and contexts of use.

Concretely, the outcome was a successful UCD process that generated a set of visualisation prototypes aligned with the needs and perceptions of urban mobility experts. Moreover, the results allowed identifying which visualisations are more adequate for knowledge extraction, depending on the underlying data and context. Those visualisations are expected to play a relevant role in decision-making and better understanding urban mobility phenomena after their future implementation. The conclusions also confirm that this UCD process for developing visualisations can produce satisfactory outcomes for the urban mobility domain.

Although visualisations are typically evaluated on an empirical basis and it is difficult to infer generalisations, the results presented herein are not limited by their validation scope, and may be used as a reference for guiding and encouraging other researchers and practitioners to take advantage of user-centred design methods for developing visualisation techniques for urban mobility and other domains.

1.2 Outline

The remainder of this thesis comprises the following topics:

- Chapter 2 – Literature Review: Presents the most relevant findings in the domains of User-Centred Design and Information Visualisation. Also, it presents the main synergies between UCD and Information Visualisation, with focus on spatiotemporal data and urban mobility. Finally, a summary of the main findings collaborates to delimiting the research context and justifying the relevance of this thesis.

[1] Website: www.move-me.mobi

- Chapter 3 – Methodology: Describes the stages for analysing datasets: data characterisation and time modelling. It also describes the issues that naturally arise when manipulating data from different sources. Finally, it describes the proposed UCD process for designing visualisations.
- Chapter 4 – Analysis of datasets: Presents the results from the analysis of datasets phase and its underlying stages.
- Chapter 5 – Results of the User-Centred design process: Presents the visualisation prototypes and other results and requirements derived from the exploratory usability tests with experts.
- Chapter 6 – Discussion: Presents the interpretation of the research findings about the UCD process and the resulting visualisation prototypes, aligned with the topics regarded in the Literature Review (Chapter 2).
- Chapter 7 – Conclusion and evolution perspectives

Chapter 2 Literature Review

This chapter comprises a critical review of the areas of User-centred Design and Information Visualisation; the latter focusing on visualising spatiotemporal data and current studies in the urban mobility domain. Finally, it links both areas by presenting the most significant UCD approaches to Information Visualisation found in current literature.

2.1 Usability and User-centred design

2.1.1 Usability

DEFINITION

A well-known definition of usability belongs to Nielsen, who describes it as “Usability has multiple components and is traditionally associated with [the] five usability attributes: learnability, efficiency, memorability, error recovery and satisfaction” (Nielsen 1993). The ISO standard 9241-11 proposes another definition of usability, defining it as “the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use” (ISO 9241-11 1998). The term product can be extended without loss of generalization to other scopes, like an interactive system.

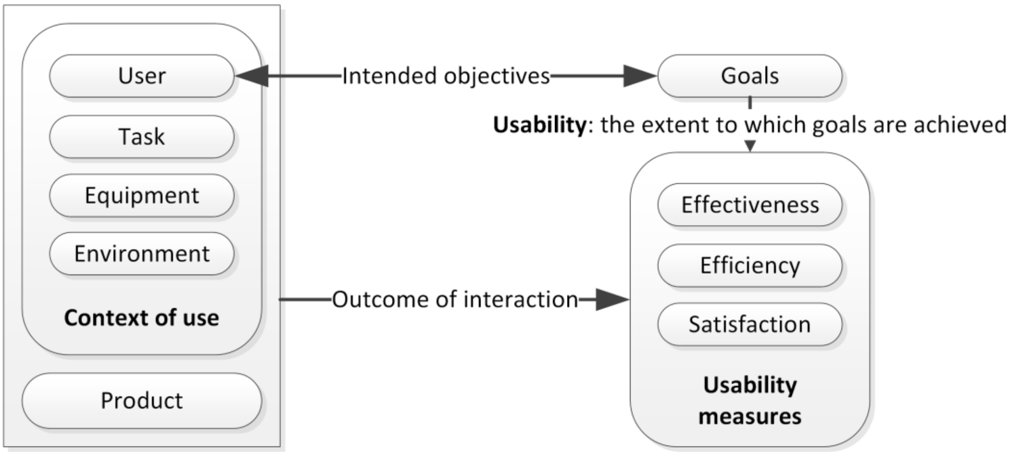


Figure 1 - Usability framework (Adapted from ISO 9241-11 1998)

CONTEXT OF USE AND GOALS

Figure 1 shows the usability framework proposed by the ISO standard. The user’s intended objectives generate the goals, i.e. what the user wants from the system. Interaction designers should understand the system’s context of use, which comprises the users themselves and the tasks they will be performing within it in order to satisfy the aforementioned goals. This includes understanding the users’ profiles, their experience and typical routines. Likewise, technological aspects are also relevant, thus implying the need of understanding the technical and environmental constraints that will surround the system use. Naturally, the users who interact with a system play a crucial role for its usability. This implies the need of understanding users and specifying their differences. Shneiderman defined three types of user abilities for specifying users: physical (age, height, keystroke speed); cognitive and perceptual (memory, learning, problem-solving, decision making) and personality (habits, attitudes towards computers, emotional states) (Shneiderman 1998).

USABILITY MEASURES

Usability measures define the extent to which a system is usable in a particular context (ISO 9241-11 1998). Therefore, they build the basis for comparing the usability between different systems. According to the aforementioned definition, the ISO standard states the following usability measures:

Effectiveness: *Determines whether users are able to perform the intended tasks or not, and up to a certain degree of completeness, achieving the respective goals related to a system.*
Efficiency: *Determines the effectiveness of users' resources and effort that are needed in order to perform the intended tasks. Typically, time is a measure for determining efficiency.*
Satisfaction: *Naturally subjective, it determines if users are comfortable with a system, and if they find it easy to use and adequate to its context of use.*

Naturally, the specification and detail of the usability measures are not subject to a general rule, thus admitting flexibility for adaptation depending on the context. However, when compared to Nielsen's definition of usability, three other measures are relevant (Nielsen 1993):

Learnability: *Determines whether a system is easy to learn or not.*
Memorability: *Determines how easy to be remembered a system is, after a considerable period of time without using it.*
Errors: *Determines if the number of errors made by users while using a system is low, and how easy is for a user to recover from error scenarios.*

Noticeably, both definitions include its respective measures naturally. All well-structured design processes should outcome systems with successful usability measures.

2.1.2 User-centred design

DEFINITION

The term User-Centred Design was introduced by Norman and Draper (Norman and Draper 1986) and has been widely acknowledged by researchers and practitioners. Moreover, other equivalent terms exist, such as Usability Engineering and Human Centred-Design. For the sake of consistency, this thesis adopts the term User-Centred Design. Despite the variety of equivalent terms and their acceptability, there is a lack of agreement upon a definition of UCD. However, several definitions were proposed throughout time (Gulliksen et al. 2005).

This review highlights two significant perspectives regarding the definition of UCD. The first emphasizes the need of having a good understanding about users, their needs and interests (Norman and Draper 1986). Nonetheless, there is no mention of involving users during the design process. The second one emphasizes the involvement of potential users of a system during the design process (Karat 1996). On the other hand, Karat considers that the lack of an agreement on a common process is also considered to be an inherent quality to UCD, giving openness for designers regarding the way they involve users along the process (Karat 1997).

The ISO 9241-210 standard formalises UCD as a process for interactive system development that focus on enhancing usability of a given system (ISO 9241-210 2008, Bevan 2009). This definition is not generally accepted, and the ISO standard does not specify the methods and techniques that should be applied throughout the design process (Jokela et al. 2003). This lack of agreement has generated efforts in order to propose a definition and a set of guiding principles for the UCD process (Gulliksen et al. 2005). However, most of the referred principles are based on usability principles found in earlier literature on usability (Gould and Lewis 1985):

1) Early focus on users and tasks: *designers must understand the users that will take benefit from a system.*

2) Empirical measurement: *users should manipulate simulations and prototypes, and their performance and reactions should be observed, recorded and analysed.*

3) Iterative design: *there is a need of having a design cycle, driven by the problems that are found during empirical measurement.*

THE UCD PROCESS

The ISO 13407 standard describes UCD as an iterative process consisting of five phases as represented in Figure 2. After planning the UCD process, the identification and further study of potential users and the contexts of use take place. This should result in the specification of user and organizational requirements, which should be constantly followed during the prototyping phase. Prototypes are finally evaluated with potential users in order to identify issues or gaps, according to users' satisfaction and the aforementioned requirements. Typically, no prototype is sufficiently valid in the first trial, which naturally implies the iterative nature of a user-centred design process.

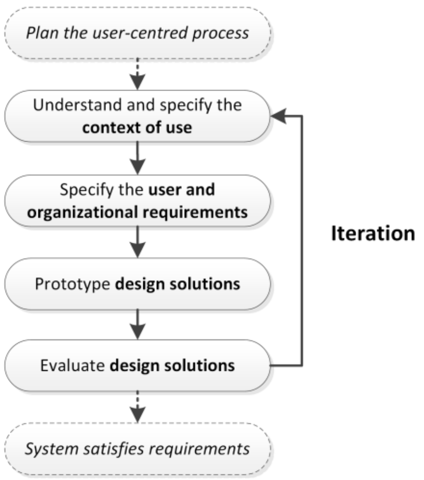


Figure 2 - User-centred design process according to ISO 9241-210 (Adapted from Jokela et. al 2003)

(i) Understand and specify the context of use

A UCD process is grounded on understanding the users of a product/system, who belong to a specific context (see Figure 1). Hence, the first step should consist on identifying its users, their tasks and the environment to which they belong. One way of specifying users is to define user groups based on characteristics or abilities (ISO 9241-210 2008, Jokela et al. 2003, Shneiderman 1998). Secondly, the identification and description of the tasks that users need a system to perform takes place. Furthermore, the description should include the overall goals to be achieved when using the system (ISO 9241-210 2008).

Finally, a thorough understanding and specification of the context of use requires the description of the social and physical environment, which includes the equipment that support the interaction between the user and a system. The ISO 9241-210 standard does not make reference to any available methods for information gathering about the context of use. Nonetheless, other works suggest widely accepted methods such as interviews, questionnaires, surveys and observation (Benyon 2010, Preece, Rogers, and Sharp 2002). The completion of this first stage of the UCD process provides the basis for eliciting requirements and elaborating the first prototypes.

(ii) Specify the user and organizational requirements

Requirements are statements about a product or system that specifies what it should do or perform (Preece, Rogers, and Sharp 2002). This second stage of the UCD process makes use of the outcomes from the first stage, structuring the retrieved information by using a variety of methods. Requirements should have a clear structure, as they might appear in different forms and levels of abstraction. Scenarios, personas and use cases are some of these methods. They are not mutually exclusive, thus a combination of them might be relevant depending on the design context.

Scenarios and personas consist on narrative descriptions that allow the exploration and discussion of contexts, needs and requirements, even though they do not describe explicitly how the user-system interaction occurs (Benyon 2010) and do not provide any flexibility for the user (Nielsen 1993). On the other hand, use cases emphasize this interaction rather than the user’s tasks (Preece, Rogers, and Sharp 2002).

The resulting requirements should be validated and probably rated according to relevance and importance for the client and users. After defining the system requirements, the design activities take place.

(iii) Prototype design solutions

The iterations of the UCD process should result in prototypes that progressively fit the users’ needs in their specific tasks. There is no problem in having simple or immature prototypes during the first iterations. Moreover, initial prototypes are crucial for validating design ideas and finding key issues that are typically easier to overcome during the initial phases, thus reducing design costs and avoiding delays (Nielsen 1993). Prototypes also answer questions and support the design team on decision-making (Preece, Rogers, and Sharp 2002). Typically, the outcomes of prototypes are paper sketches, mock-ups or simulations. The iteration of the prototyping phase should continue until the desired design goals are met.

In order to save time and costs and still be able to develop solutions that can be tested by users, designers can either limit the number of features in the prototype or reduce the level of functionality of the features in a way that they apparently work but do not actually do anything (Nielsen 1993). These two perspectives on prototyping are characterized as vertical and horizontal prototyping respectively (see Figure 3).

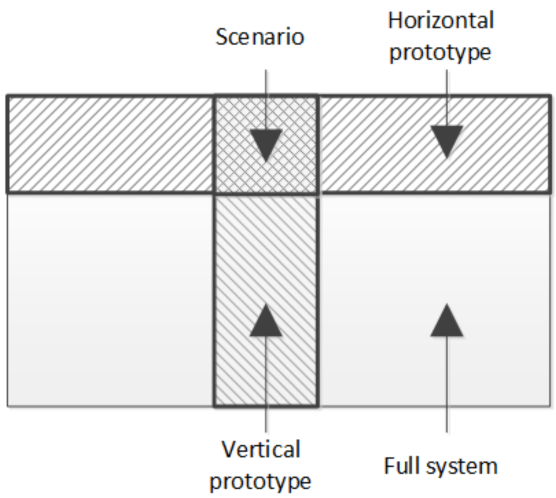


Figure 3 - The two dimensions of prototyping (Adapted from Nielsen 1993)

A horizontal prototype is a representation (simulation) of the interface where no real activities can be performed. Instead, it contains the entire user interface, so users should be able to navigate and explore the system but without retrieving any real results from these commands. A vertical prototype consists of a limited part of the full system, which can be tested under real conditions with real user tasks. Therefore, it is possible to retrieve real results within their limitations.

(iv) Evaluate design solutions

After producing a set of prototypes, the evaluation of the design solutions takes place. The goal of this stage is to generate feedback from the client and users in order to fix detected issues or inconsistencies and improve the system, ensuring that the proposed design fulfils the user and organizational requirements and usability goals (Benyon 2010). Evaluation can employ usability experts or regular users, through expert or participant-based methods respectively.

Expert-based methods employ usability experts during the evaluation of prototypes. Examples of those methods are: heuristic evaluations, cognitive walkthroughs and consistency inspections (Benyon 2010, Nielsen 1993). Participant-based methods employ potential system users. Usability tests are one of the essential methods for evaluating usability, as real system users manipulate the system. Basically, a test user performs a set of pre-defined tasks within the system interface. These tests might occur in pre-set laboratories or in real work environments, and are always carried by a moderator (Nielsen 1993).

CONCURRENT UCD APPROACHES

The ISO 9241-210 cannot be regarded as a general rule. Moreover, it can be considered as an effort for the development of guidelines for researchers and practitioners, even though it resembles UCD as a highly sequential process, which is frequently different from reality. The phases that compose the ISO standard are based on previous approaches to the UCD process. A direct comparison between methods shows that those phases are typically inherent to the existing approaches, only differing by the way they are ordered or performed along the processes.

(i) Waterfall model

The waterfall model was the dominant approach for decades in the software development industry. It is characterised by a rigid structure oriented to projects where schedule commitment is crucial. The result of this structure is an inflexible method but typically easy to track, given that deadlines for each phase are established in the beginning of the process. Nonetheless, there is limited feedback among its phases, especially between non-adjacent ones. Typically, this might cause the late discovery of issues in the design process. As a consequence, if the issues are critical in the context of the project, the process might need to restart at an earlier phase and go through the subsequent ones, thus leading to schedule slippage and costs overruns.

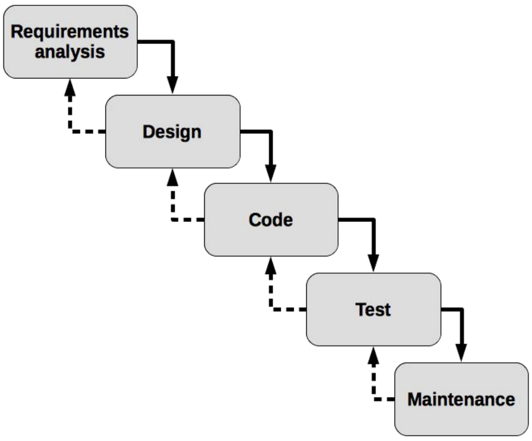


Figure 4 - A waterfall model approach (Adapted from Preece, Rogers and Sharp 2002)

(ii) Spiral model

The spiral model proposed by Boehm is an effort to overcome the limitations of the waterfall model, by bringing flexibility and enhanced feedback between all project phases through a highly iterative process. In addition, it offers a risk-driven approach to the design process that encourages the consideration of design alternatives. Concretely, the spiral model consists of four main phases: 1) Definition of objectives, alternatives and constraints; 2) Risk analysis and Prototyping; 3) Developing and Testing and 4) Planning the next phase. The underlying concept of this model relies on minimising risk by making repeated use of prototypes in every iteration until a prototype is acknowledged as final. Noticeably, phases 2 and 4 concentrates on answering to two questions that are inherent to this model: “What should be done next?” and “For how long should it continue?” (Boehm 1988). This approach is costly but suitable for high-risk and complex projects.

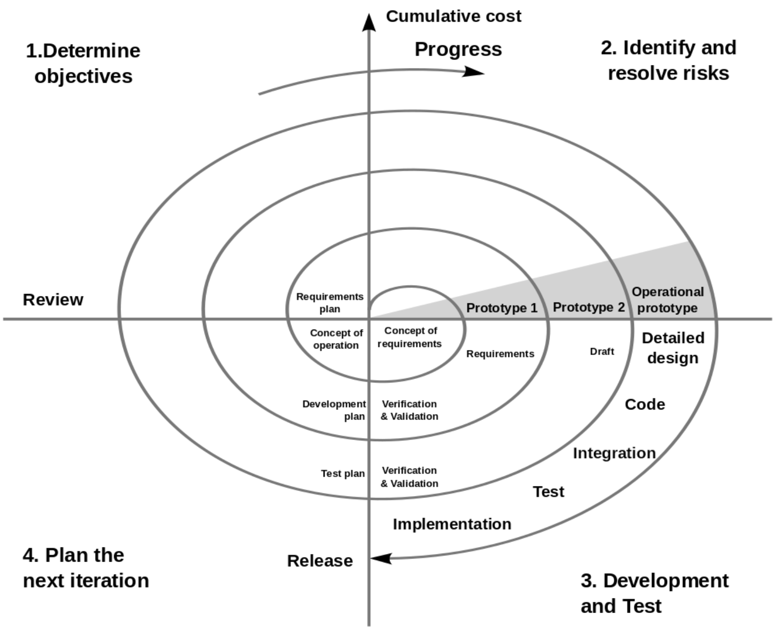


Figure 5 - A spiral model approach (Adapted from Boehm 1988)

(iii) Evolutionary model

The Star lifecycle model derives from empirical studies with interface designers and the on-going projects (Hartson and Hix 1989). This approach considers evaluation in the centre of design activities. At each phase of the design process, evaluation should take place with users. This contrasts with the spiral model, in the sense that smaller loops take place within the star model. In addition, there is a minimized stress on the ordering constraints of the design process. It is possible to start (or move) to any design phase, assuming that the results of the preceding phase were already evaluated. For example, developers are not constrained from prototyping only after requirements are established.

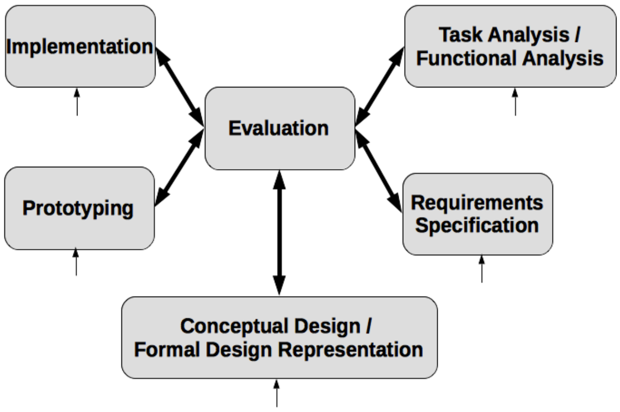


Figure 6 - The star lifecycle model approach (Adapted from Hartson and Hix 1989)

EXPECTED BENEFITS OF UCD

A well-structured design process following a user-centred approach shall offer substantial benefits to the usability of a system, as well as financial advantages for the system developer (Earthy 1998):

- Increased productivity and quality: *Systems with good usability can allow an increase on work output with a reduced occurrence of errors.*
- Reduction of production costs: *A UCD approach can prevent the over-design of a system, thus reducing overall development times and costs. It facilitates the discovery of issues during the early development phases.*
- Reduced of training and support costs: *Systems with good usability require less training and user support, thus reducing their related costs.*

2.2 Information Visualisation

Information Visualisation is a recent yet wide field that has been gaining attention from researchers and practitioners. It consists of a multi-disciplinary area that intersects fields such as Computer Science, Human-Computer Interaction, Design and Cognitive Psychology. The maturation of the Information Age and the settlement of the Big Data paradigm motivate the need of visually representing data from large datasets to support knowledge extraction. Visualisation has a crucial role in human cognitive systems, especially for the fact that visual displays provide the highest bandwidth channel from the computer to the human (Ware 2004).

DEFINITION

One of the first definitions of Information Visualisation regards it as “the use of computer-supported interactive visual representations of abstract data to amplify cognition” (Shneiderman, Mackinlay, and Card 1999). Spence defines it as the formation of a mental model or mental image of data (Spence 2007). Despite the diversity of applications inside itself, it is a consensus that visualisation techniques can help informing and improve analysis and decision-making. Furthermore, Ward *et al.* emphasize that Information Visualisation does not rely solely on computer graphics, and that it is the connection to data that is crucial (Ward, Grinstein, and Keim 2010).

Information Visualisation is one of the major areas of Visualisation besides Scientific Visualisation and Geovisualisation. Even though this classification is generally acknowledged, this classification lacks precision and it is possible to identify substantial overlap between areas. Information Visualisation generally concerns the representation of abstract data like prices or social habits. Scientific Visualisation typically comprises scientific data which might include a spatial component like three-dimensional simulations (see Figure 7). Geovisualisation is considered to be similar to Scientific Visualisation, differing by having maps as the centre of data display. The scope of this thesis naturally admits the overlap (or a better term: synergy) between Information Visualisation and Geovisualisation, given that abstract information about urban mobility is embedded into a spatiotemporal context but it is not necessarily restricted to geographic representations.

A common misconception consists in overlapping the definitions of Information Visualisation and Visualisation tools. The latter can be defined as computer-based systems which are designed to display encoded data with a view in order to support the visualisation process (Spence 2007).

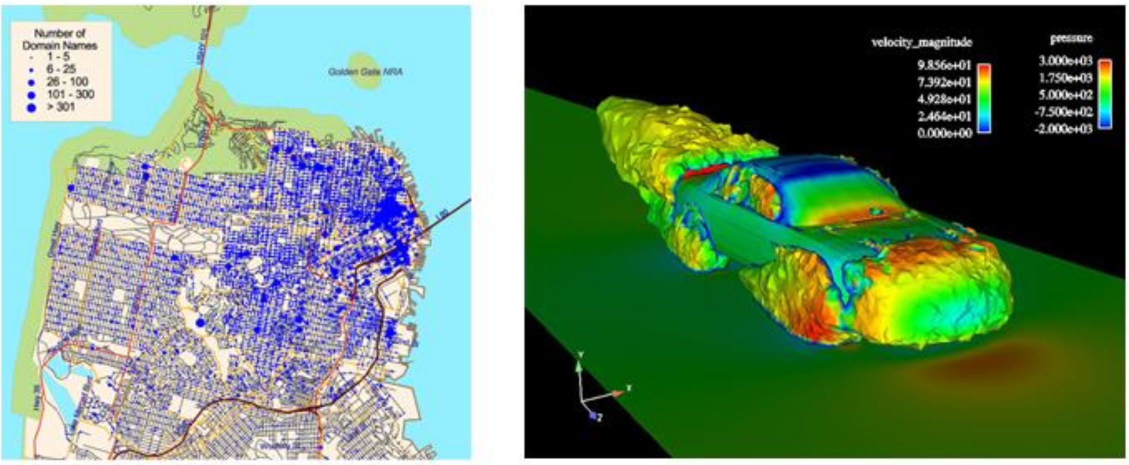


Figure 7 - An example of information visualisation (left) in contrast to a scientific visualisation produced by scientific computing application (right) (Sources: http://www.infovis-wiki.net and http://www.designworldonline.com)

TAXONOMIES FOR DATA TYPES AND VISUALISATION TECHNIQUES

Visualisations are not merely based on data; they are true representations of it. Therefore, it is crucial to analyse and classify the data concerned before starting the development of visualisation techniques. Shneiderman proposed a taxonomy for seven types of data (Shneiderman 1996):

- One-dimensional (1D): e.g. alphabetic lists of names
- Two-dimensional (2D): e.g. geographic maps
- Three-dimensional (3D): e.g. molecules
- N-dimensional (ND): e.g. relational databases
- Temporal: e.g. timelines associated to medical records
- Tree: e.g. sales data
- Network: e.g. links between pages in the World Wide Web

The aforementioned taxonomy is not stiff and overlaps are possible between two or more data types. The main goal of providing taxonomy for data is to facilitate discussion and lead to discoveries that might be useful for developing visualisations. Prior to developing visualisations, it is important to ensure that data is retrieved from a reliable and accurate source (Ward, Grinstein, and Keim 2010). Shneiderman suggests a useful starting point for designing visualisations named "visual information seeking mantra": "Overview first, zoom and filter, then details-on-demand" (Shneiderman 1996):

- Overview: *gain overview about the object in analysis*
- Zoom: *zoom in on objects that are of interest*
- Filter: *filter unwanted objects*
- Details-on-demand: *select an object (or a group of) and get further details*
- Relate: *view relationships between objects*
- History: *keep track of actions to support undo, replay and progressive refinement*
- Extract: *allow extraction of subgroups of objects and of query parameters*

The pluralistic nature of data implies a variety of visualisation techniques to represent it. Ward *et al.* divide them into six broad categories: spatial data; geospatial data; multivariate data; trees/graphs/networks and text/documents (Ward, Grinstein, and Keim 2010).

VISUALISATION AND INTERACTION

Technology has made responsive interaction possible within the context of visualisation. Visualisation tools now benefit from the capability of facilitating users to change their view about data. This change of view can be defined as movement within the information space (Spence 2007). Rather than static visual representations of information, visualisation tools allow users to interact with it through navigation and have immediate feedback. The example in Figure 8 depicts a highly responsive visualisation tool that provides an initial overview of the density of immigration settlement of different foreign-groups across the United States. It allows further exploration by filtering for specific nationalities, skimming through a timeline or zooming for details, in accordance to what has been proposed by Shneiderman.

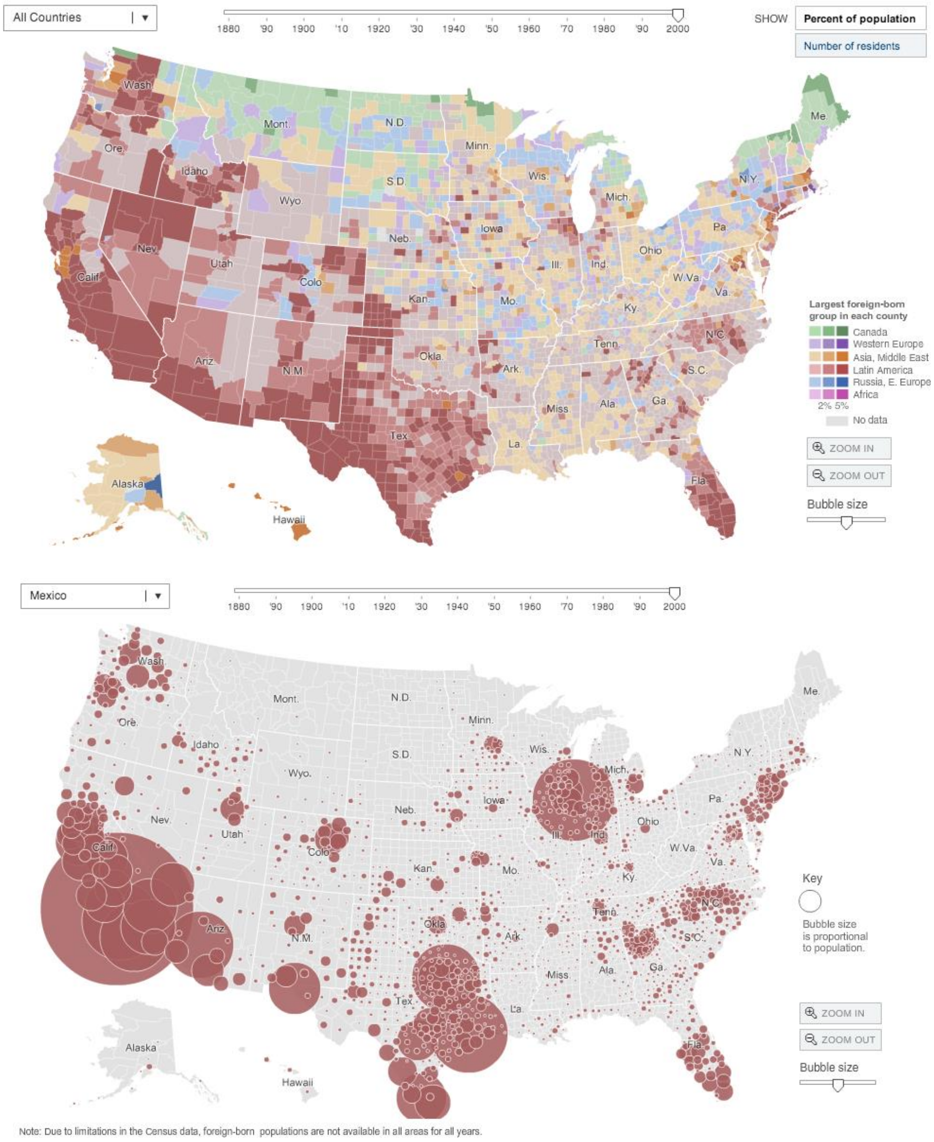


Figure 8 - Immigration settlement across the USA. Overview of immigration by different nationalities is shown on the top. Filtering by Mexican citizens immediately provides the visualisation on the bottom (Source: http://www.nytimes.com/interactive/2009/03/10/us/20090310-immigration-explorer.html)

VISUALISATION OF SPATIOTEMPORAL DATA

Depending on the context of application, visualising spatiotemporal data can also make use of non map-based techniques, thus adopting abstract graphic entities such as lines, bars, circles or other geometric forms. However, when the spatial component is of visual concern by the users who will take benefit from visualisations, approaches take place in order to suggest ways of presenting both dimensions simultaneously.

Two main visualisation approaches have been applied for visualising spatiotemporal data. The first attempt appears within the context of Time Geography coined by Hägerstraand, and consists of representing paths of moving instances through a time dimension which is perpendicular to a geographic plane – which is defined as a Space-Time-Cube (Hägerstraand 1970). Given that time is an inherent component of the visualisation, it is possible to continuously track the movement of an entity through space over a timespan. However, the amount of entities to be plotted might imply intense visual cluttering, which would require workarounds in order to reduce it. Hägerstraand's approach precedes the era of Geographic Information Systems (GIS), even though it is possible to identify applications of the Space-Time-Cube approach to GIS (see Figure 9 left) (Kraak and Koussoulakou 2005).

The second approach consists of using animated maps (see Figure 9 right) to allow users to visualise changes in spatial data over time. At every frame of the animation, the position of an entity is redrawn. In order to avoid visual clutter, animated maps should preferably plot spatiotemporal data on a static bi-dimensional map. However, this requires additional measures in order to avoid designing visualisations that force users to create mental associations between different time intervals. Instead, they should allow instant view of changes of an entity through time. Methods for displaying data within animated maps were proposed by Adrienko *et al.* (Adrienko, Adrienko, and Gatalsky 2005).

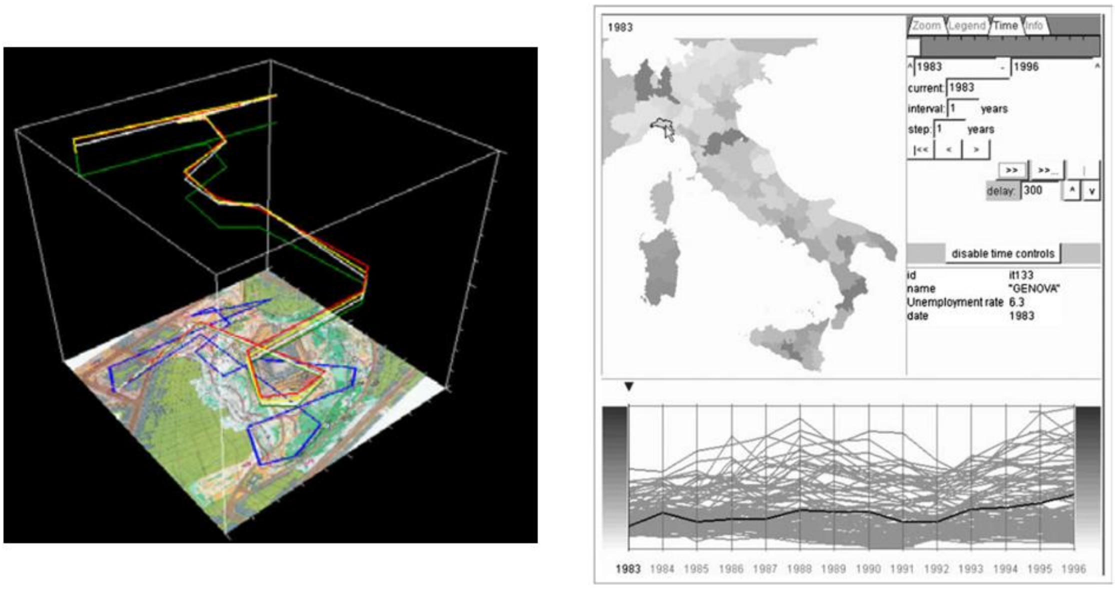


Figure 9 - Differences of representation between a space-time-cube (left) and an animated map representing spatiotemporal data over a static 2D map (right)

Sources: (Adrienko, Adrienko, and Gatalsky 2005, Kraak and Koussoulakou 2005)

VISUALISATIONS FOR URBAN MOBILITY

Urban mobility data is typically collected from systems such as Global Positioning System (GPS) or telematics systems such as Automatic Vehicle Location (AVL) and Automatic Passenger Count (APC). There are uncountable examples of visualisations of this kind available online, although they typically resemble more of artistic experiments (see Figure 10) than systematic approaches with actual intentions of collaborating with visual analytics for knowledge extraction.

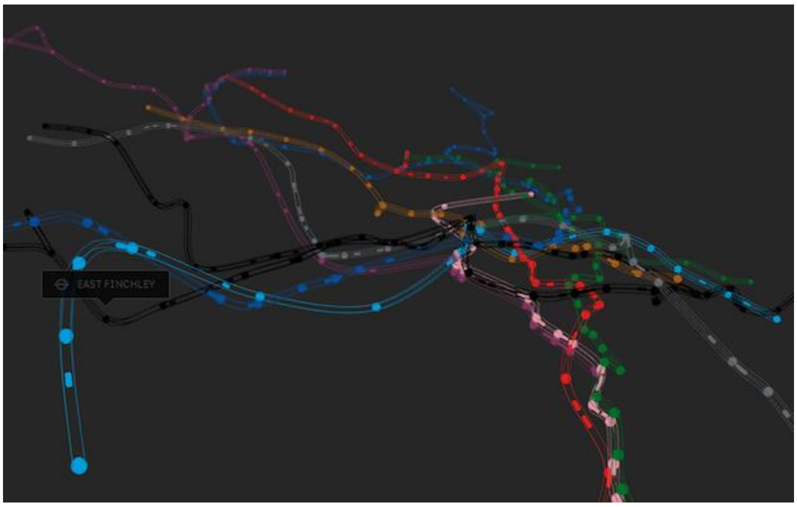


Figure 10 – Artistic visualisation experiment of the London Underground network

Source: http://brunoimbrizi.com/experiments/#/07

We present recent significant studies that explore visualisation of urban mobility. They rely either on the visual representation of information, or on the presentation of new techniques for extracting knowledge with the support of visual analytic tools, or on the algorithmic aspects for developing visualisations.

MetroViz (Du et al. 2013) is a visualisation tool for supporting the exploration of historical data of a bus network. In accordance to Shneiderman's recommendations, *MetroViz* follows an Overview-to-Detail approach, progressively revealing details according to users' input. It features an overview of the bus transport network by laying all stops on a map. Further information is made available by clicking on each stop, allowing detailed information about average patronage and individual information for each bus or route. Users participate during the evaluation phases through usability tests by executing pre-determined tasks.

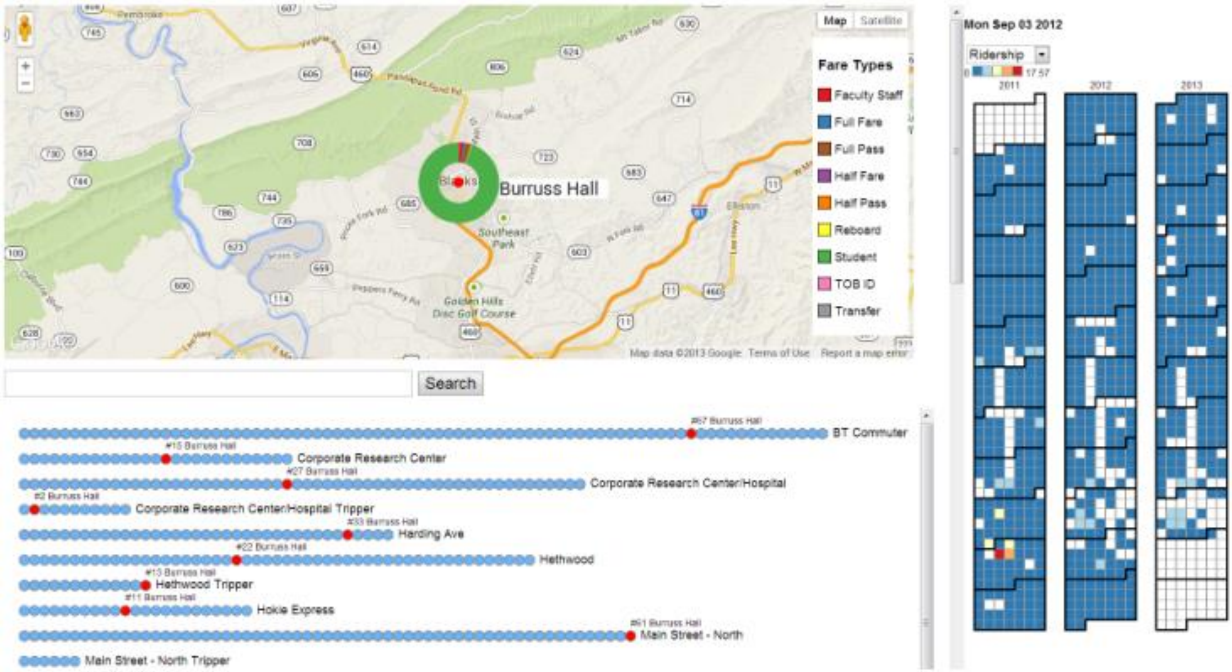


Figure 11 – Metro Viz visualisation tool (Source: Du et al. 2013)

Two studies took advantage of mobile network traffic for understanding urban mobility trends. The first one focuses on the use of this kind of data for identifying social patterns and questions whether visual representations could be used successfully in order to detect them (Sagl, Loidl, and Beinat 2012). The second presents *AllAboard*[2] (see Figure 12), a visual analytics tool for optimizing and exploring public transport networks by using mobile network traffic data (Lorenzo et al. 2014).

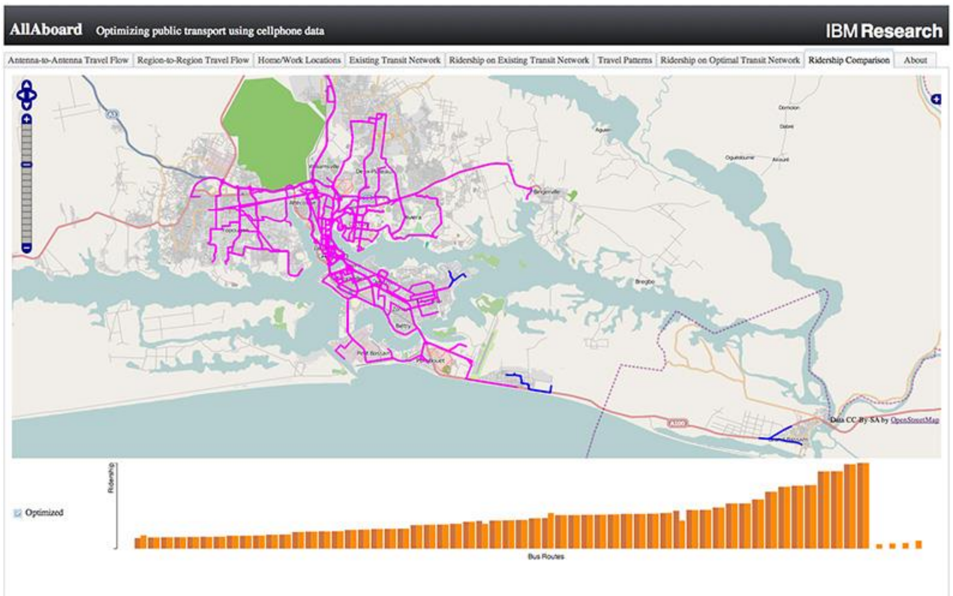


Figure 12 - AllAboard user interface (Source: footnote 2)

[2] Website: http://researcher.watson.ibm.com/researcher/view_group_subpage.php?id=4746

Finally, another study carried by Polisciuc *et al.* investigates anomalies on the number of passengers for a bus network (Polisciuc et al. 2013). It does so by developing and experimenting visualisation techniques that benefit from computer graphics techniques (see Figure 13). The visualisation represents the standard deviation of the number of passengers for each bus stop in Coimbra during the day. Red spots indicate an abnormal increase of passengers at a stop (positive deviation). Green spots represent negative deviation.

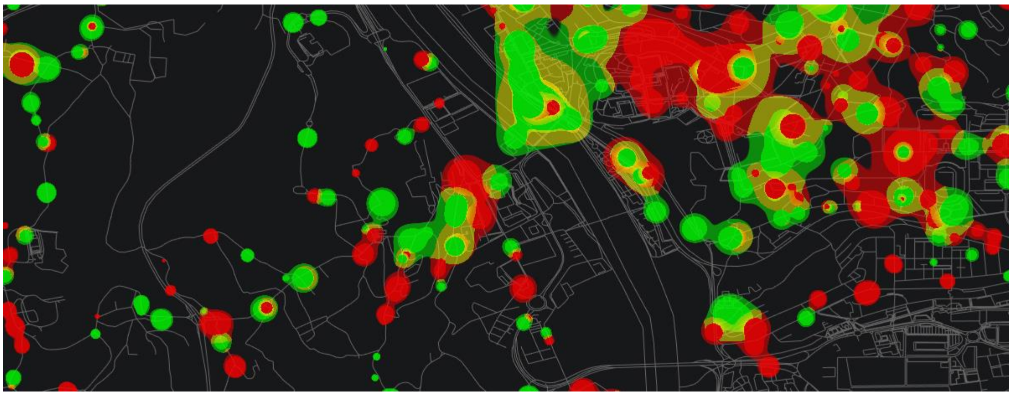


Figure 13 - Visualisation of urban mobility data (Source: http://fmachado.dei.uc.pt/sandbox/visualizing-urban-mobility)

2.3 User-centred design approaches to Information Visualisation

The involvement of users during the whole development process of visualisations is well acknowledged by researchers (Christopher et al. 2003, Slocum et al. 2003, Robinson et al. 2005, Tory and Moller 2005, Roth et al. 2010, Lian Chee et al. 2011). Visualisations have mostly been technology-driven, and it is now possible to identify a shift towards more user-centred approaches. These approaches differ on the way users are engaged throughout the process. These methods appear in diverse contexts such as epidemiology, hydrography and crime spotting visualisations, although they can be analysed from a general UCD perspective for visualisation. To the best of our knowledge, no traces of UCD approaches to visualisations for urban mobility context were found. Hence, this section exhibits the most significant results from other areas.

Slocum proposes a UCD process consisting on a six-phase cascade for the creation of a visualisation tool for water-balance issues: 1) prototyping; 2) domain expert evaluation; 3) software refinement; 4) usability expert evaluation; 5) software refinement and 6) decision maker evaluation (Slocum et al. 2003). A noticeable disadvantage is that end-users only participate in the end of the design process, after domain and usability experts have addressed the design and functionality related issues.

The method proposed by Robinson *et al.* extensively involves users throughout the process along six phases: 1) work domain analysis; 2) conceptual development; 3) prototyping; 4) interaction and usability studies; 5) implementation and 6) debugging (Robinson et al. 2005). Phases 2 to 5 occur multiple times in a loop. The work domain analysis phase consists of the first contact between stakeholders and developers, in which they communicate the initial ideas and requirements. The conceptual development outlines the results from the work domain study. This approach suggests that if the work domain is not completely understood, the prototyping phase shall not proceed. From a critical perspective, it is still possible to proceed with prototyping, looping back to the work domain analysis phase if needed.

In response to this issue, Roth *et al.* proposes a modification to Robinson *et al.*'s UCD process with the following phases: 1) prototyping; 2) interaction and usability studies; 3) work domain analysis; 4) conceptual development; 5) implementation and 6) debugging (Roth et al. 2010). Prototyping now has an initial role in the process, where designers develop visualisations according to the way they think about them. After performing interaction and usability studies, the prototype is used as a final component of the work domain analysis to catch new ideas from users. Finally, the results from the interaction and usability studies and work domain analysis formalise the conceptual development phase.

It might happen that the potential users of visualisation tools are not familiar with visualisation concepts, or are not aware of their true potential. This might compromise the effectiveness of their participation. In order to overcome these issues, Lian Chee *et al.* (Lian Chee et al. 2011) propose relevant additional phases based on the approaches of Robinson *et al.* and Roth *et al* (see Figure 14). These additional phases improve awareness of users about the power of visualisation in a general context, but also within their own context by visualising their own data.

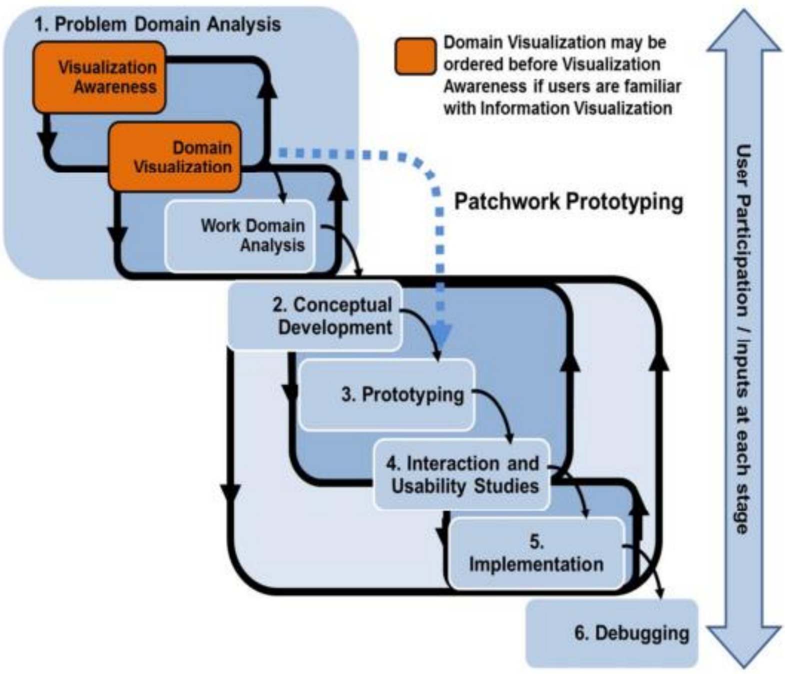


Figure 14 - UCD approach for visualisations proposed by Lian Chee *et al.* (Source: Lian Chee *et al.* 2011)

The Visualisation Awareness phase consists of introducing general concepts of Information Visualisation to users in case they are not familiar with the field. In this phase, the first discussions for retrieving ideas of the development of their own visualisations are held. The Domain Visualisation phase incorporates the development of the first prototypes tailored to the users' own data.

THE ROLE OF EVALUATION

Visualisation techniques and tools abound, but it is still possible to identify lack of engagement of visualisation researchers with users. In fact, visualisation tools are useless if users cannot effectively interact with them, thus naturally compromising the processes of knowledge extraction and decision-making. A study by Ellis and Dix – one of the few about evaluation for Information Visualisation – supports this statement: after analysing 65 papers describing new visualisations, it finds that only 12 of them engaged users within evaluation processes (Ellis and Dix 2006). Moreover, the effectiveness of the evaluation is also questioned, given that a small amount of studies that implemented evaluation as part of the development process are somewhat problematic. A further complicating factor is the inherent nature of visualisation itself that turns evaluation into a complicated process. Again, the authors indicate some factors that contribute to making evaluation in Information Visualisation a hard process:

1) Variety of datasets: *despite some earlier efforts in creating a standard for datasets, datasets are heterogeneous and hardens the evaluation process, as it might limit the availability and quality of data to be visualised.*
2) Indeterminacy of tasks: *the tasks to be performed during the evaluation process are usually more structured, which usually differs from the ones to be performed in "real life" that are more exploratory; these ones are harder to replicate in an experiment*
3) Participants in context: *depending on the complexity of the application context, participants need to have a clear understanding of the problem that the visualisations are trying to solve. Some authors suggest that it is possible to obtain better information by involving domain or usability experts, even though it is typically harder to have access to those people (Christopher et al. 2003, Tory and Moller 2005).*

2.4 Conclusions and research contributions

User-centred design approaches are widely accepted and well posed. Concurrent approaches towards involving users in the design of interactive systems enforce the fact that it is difficult to engage users in order to design usable and useful systems. Moreover, it shows that there is not a general rule for carrying a UCD process. However, it is crucial to formalise it by developing a coherent structure that matches the context of application into which is embedded.

It is evident that current UCD approaches are highly iterative. In the context of Information Visualisation, prototyping during initial phases seems to be promising, rather than only on later stages. This contrasts with the ISO 9241-210 standard, which places the prototyping phase after the specification of user and organizational requirements. Early prototyping does not deviate the design process from the user-centred perspective at all, as it is possible to iterate and refine prototypes by revisiting the work domain analysis while it is needed. Likewise, early prototypes can benefit domain users if they are not familiar with Information Visualisation.

We presented recent and significant studies on visualisations for urban mobility data, each of them belonging to different facets. As mentioned in this chapter, Information Visualisation is an interdisciplinary field, thus it is expectable (and positive) to find studies with diverse perspectives within this field. Nonetheless, the existence of few relevant studies on user-centred design approaches – especially within the context of urban mobility domain – was considered of concern.

Based on our findings during the literature review, we conclude that this research offers a novel contribution by carrying a thorough user-centred design approach for the development of visualisations in the urban mobility domain, based on similar studies applied to the development of visualisations within other domains.

Chapter 3 Methodology

This research addresses the problem of developing visualisations for urban mobility. Our focus on adopting a user-centred approach is based on the lack of thorough user engagement during the development of visualisations in current literature, especially in the domain of urban mobility, although it is possible to identify UCD approaches that were successfully applied in other domains as mentioned in Section 2.3. Besides the visualisations themselves, we are also interested in assessing the benefits of involving users during the design process.

Our research methodology consisted of two phases: analysis of urban mobility datasets and a User-Centred Design process for developing visualisations. The analysis of urban mobility datasets consisted of two activities: *exploration of datasets*, which identified the data primitives they contained and the inherent issues that arise when manipulating raw data; a formal *data characterisation* and *time modelling*, in order to understand how data and time should be regarded within each visualisation. Finally, the UCD process took advantage of the results of the first phase for the development of visualisation prototypes.

We collected three datasets acquired from different sources that summed up to 3 gigabytes of raw data. The first contained users' requests from Move-me, a mobile application that provides real-time information about public transport. The second contained bus-ticketing data from Sociedade de Transportes Coletivos do Porto S.A. (STCP). The latter contained readings from traffic counter sensors for a set of streets from Porto. Data primitives were conceptually tied to a date, time and location. Depending on the data primitive, it was necessary to translate the conceptual locations (e.g. "Aeroporto") into geographical coordinates, which is commonly regarded as *geocoding*. Finally, raw data were further ported to a Database Management System (DBMS).

3.1 Analysis of datasets

3.1.1 Exploration of datasets

The exploration phase analysed the contents of each dataset, in order to identify which data primitives they contained, their data type and format. Datasets are typically heterogeneous in terms of structure, as they derive from different sources (Ellis and Dix 2006). This creates technological constraints that should be identified prior to the design stage, so as to prevent unforeseen drawbacks. Most of these constraints were related to the format of data primitives (e.g. special characters at the end of a string, data in XML syntax), which usually required manual intervention before parsing them to a DBMS. Four questions guided this exploration:

- What is the nature of each dataset?
- What is the meaning of each entry (row) in each dataset?
- Which attributes (columns) are relevant as inputs for our visualisations?
- Does the relevant attributes require prior adaptation before using them as inputs for our visualisations?

3.1.2 Data characterisation

In order to provide a formal characterisation of data, this research considered the *pyramid framework* (see Figure 15) proposed by Mennis *et al.*, which defines three perspectives for data: *location* (where is it?), *time* (when is it?) and *theme* (what is it made of?) (Mennis, Peuquet, and Qian 2000). These perspectives form the *data component*, which can be regarded as un-interpreted observational data. The derived knowledge from those perspectives forms the *knowledge component*, and can be regarded a semantic object: a conceptual entity. A semantic object may belong to a taxonomy (classification) and may have part-whole relationships with other semantic objects (partonomy).

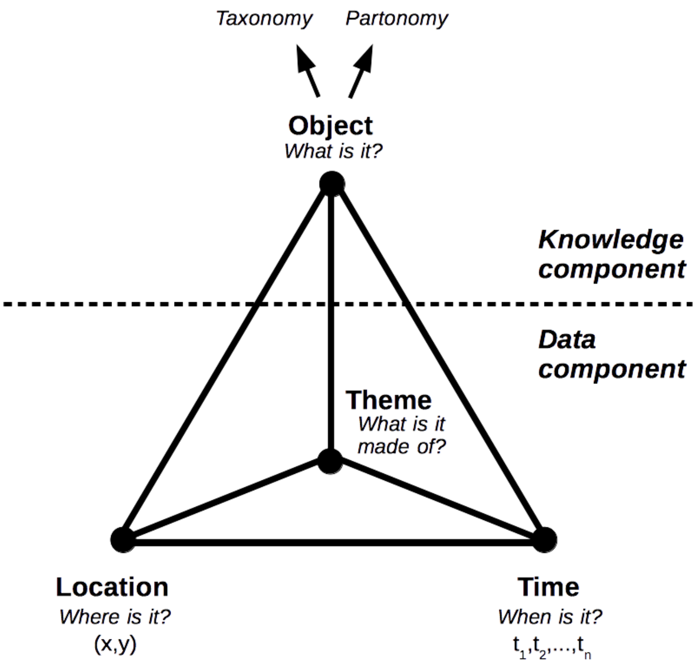


Figure 15 - Pyramid framework for characterising data (Adapted from Mennis, Peuquet, and Qian 2000)

3.1.3 Time modelling for visualisations

We considered a formal time modelling for designing a time model that fit adequately to the characteristics of our visualisations, i.e. how time should be conceived while querying data to be used as inputs for visualisations. According to Aigner *et al.*, it is possible to model time using a variety of models or taxonomies. That depends on the context in which time is being modelled (Aigner et al. 2011). The adopted time model comprises four perspectives: *scale*, *scope*, *arrangement* and *viewpoint*, which are briefly explained herein:

Scale regards the distribution of events along the time domain, which can be *ordinal*, *discrete* and *continuous*. An ordinal time domain only considers the relative order of the events and can be useful when no quantitative information about time is available; i.e. given two events, it is only possible to discern and convey meaningful information if both are considered at the same time. A discrete time domain maps the time of an event to a timestamp and is frequently used in information systems. Finally, a continuous time domain maps time to the set of real numbers, thus being possible to exist an event between any two events given.

Scope regards the quantification of events, which can be *point-based* or *interval-based*. Point-based domains include events with null temporal extent, i.e. it is possible to know when they happened but not for how long. Interval-based domains overcome this "limitation" by grouping events in order to introduce the notion of temporal extent.

Arrangement regards how time domain is arranged, which can be *linear* or *cyclic*. Linear domains resemble our natural perception of time, where events are lined from the earliest to the farthest one. Cyclic domains group events according to reoccurring time cycles like seasons or weekdays and weekends.

Viewpoint regards the perspective at which time is viewed, which can be *ordered*, *branched* or *multi-perspective*. An ordered viewpoint aligns events sequentially according to an attribute. A branched viewpoint considers alternative scenarios for a sequence of events that will happen during a certain timespan, which is useful in decision-making processes and simulations. Finally, multi-perspective considers events that belong to distinct time instants depending on how they are interpreted (e.g. a birth of a child is tied to a date, but the date in which he/she was registered might differ).

3.2 A User-Centred Design process for developing visualisations

We considered the UCD methodology proposed by Lian Chee *et al.* described in Section 2.3 with some changes (see Figure 16). This choice enforces the commitment of involving potential users throughout all phases and acknowledges rapid prototyping in order to foster feedback and to overcome time constraints. According to Section 2.3, prototyping could be performed anytime, even if not all specifications or requirements were gathered (Lian Chee et al. 2011).

The aforementioned changes included a preliminary phase that consisted in selecting experts in urban mobility interested in participating in this research and studying the underlying context of use for visualisations. Implementation and Debugging phases were removed, as the objective did not consist in developing a complete system for visualising urban mobility data, but identifying which visualisations were more adequate for representing the urban mobility data we collected. Therefore, prototyping and evaluation were the core of this research.

We conducted semi-structured interviews with experts for evaluating visualisations and retrieving feedback and ideas. Quantitative approaches such as questionnaires were not adopted, as they would not outcome detailed results, especially when considering the exploratory nature of this research. Moreover, it is well acknowledged that interviews best suit exploratory studies (Nielsen 1993). Most of the sessions with experts took place at their working environments. Given that most experts already had knowledge about Information Visualisation, the Visualisation Awareness phase did not receive extensive focus. All interviews were recorded in audio format and had the support of note taking, in order to ensure that all qualitative inputs were considered during our analysis.

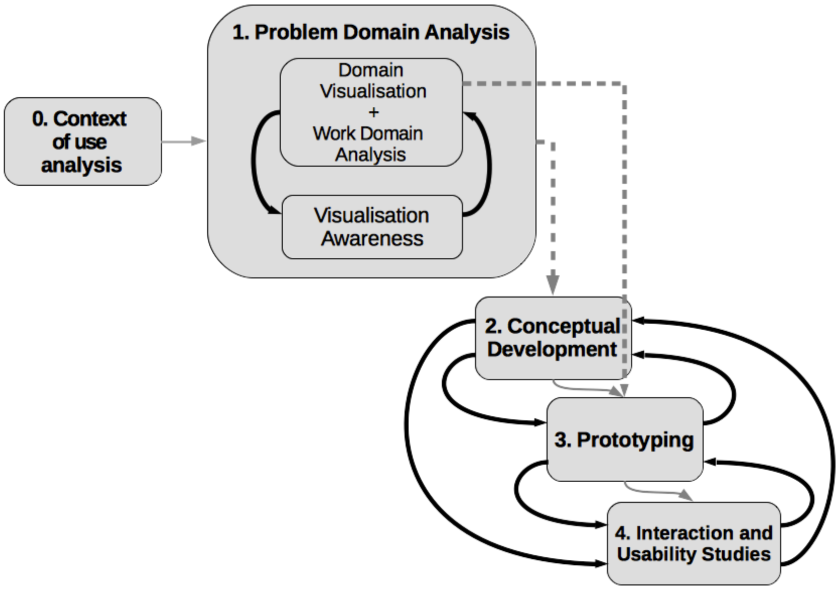


Figure 16 - UCD process used in this project adapted from Lian Chee *et al.*

3.2.1 Context of use analysis

A typical User-Centred Design cycle starts when a stakeholder requests the design of a system. Since this research went the opposite way, it was crucial to identify experts willing to participate on this process, and to make effective use of visualisations in the future. 10 experts in urban mobility from Porto participated in this research:

- 5 experts from strategic and operational staff levels of transport organisations
- 2 experts from OPT – Optimização e Planeamento de Transporte S.A, which develops solutions for transport planning, management and public information;
- 2 experts from the Urban Mobility Division of Porto City Hall
- 1 researcher on urban mobility from the Faculty of Engineering of University of Porto

3.2.2 Problem Domain Analysis

We developed vertical prototypes using real data and limited functionality. For the first exploratory usability test with experts, we proposed an initial set of prototypes in order to catalyse the generation of ideas for the further iterations of the UCD cycle. Also, given the limited available time of participants, we wanted to retrieve substantial feedback in every contact opportunity with them.

The first iteration of Domain Visualisation consisted in presenting static mock-ups and the first prototypes, which showcased real data from specific periods of time and events (e.g. St. John's Night in Porto). Experts also had the opportunity to manipulate the prototypes. The subsequent iterations of this phase were named 'Interaction and Usability Studies', which instead presented more advanced prototypes.

Work Domain Analysis phase was integrated with Domain Visualisation and Interaction/Usability Studies phase, and consisted in gathering requirements and potential suggestions for visualisations based on their context of use.

Visualisation Awareness consisted in the presentation of not more than 10 examples of visualisations showcasing different techniques, not necessarily related to the urban mobility domain. The goal was to suggest additional opportunities to be explored by demonstrating their application in other domains.

3.2.3 Conceptual Development

This phase was re-visited after the exploratory usability tests with experts during the Interaction and Usability studies. We analysed the recorded audio files from the interviews, which included requirements, issues found and ideas. The proceeding conclusions were considered on further iterations of Prototyping phase.

3.2.4 Prototyping

As we expected to retrieve the most possible feedback from experts within our planned timeframe and to investigate the potential of our visualisation prototypes, we were concerned with developing vertical prototypes to be evaluated during the meetings. We wanted that experts could actually interact with visualisations by zooming, querying details on demand or filtering data according to pre-defined parameters. This naturally implied the risk of investing significant time in prototyping complex visualisations that could be of no practical use by urban mobility experts if they were not satisfied with them.

Visualisations were classified into two frames or reference: geographic and abstract (non map-based). Geographic visualisations were built with Google Maps API, Google Earth and Gephi[3]; Abstract visualisations were built with HTML5 and D3 (Data-Driven Documents)[4]. These technologies are capable of handling large volumes of data, producing appealing visualisations. As this research is concerned with the inherent aspects of visual representations of data, code optimisation or rendering times were not prioritised. Experts were aware that prototypes could be eventually slow while manipulating them or querying data for long time extents.

3.2.5 Interaction and Usability Studies with experts

Considering our aim of assessing how useful our prototypes could be, including to which extent they could be improved, changed or even discarded, we invited experts referred in Section 3.2.1 for participating on individual or group sessions of approximately 90 minutes long, depending on their availability. Each exploratory usability test applied a semi-structured interview (see APPENDIX B: Guide for semi-structured interview (Domain Visualisation phase)) that was checked and refined prior to its execution during pilot tests with non-experts. In our case, pilot tests did not necessarily involve people with experience in urban mobility, although this does not compromise the validity of these tests. Moreover, users that are not familiar with the theme can also detect issues during the experimental design of the tests (Nielsen 1993).

---

[3] Website: gephi.org

[4] Website: d3js.org

The exploratory usability tests with experts followed the *coaching method* proposed by Mack and Robinson (Mack and Robinson 1992) combined with a semi-structured interview given during the session. The combination of different usability methods is well acknowledged in usability research (Nielsen 1993). The most distinguishing feature of the coaching method in comparison to other evaluation methods relies on the active involvement of the mediator during the test, who can actually steer users in the right direction while using a system. The coaching method contrasts with other techniques such as think-aloud, where the mediator's involvement should be kept to a minimum level. Considering the exploratory nature of our study, the coaching method was considered more adequate.

Exploratory usability tests consisted of four phases:

- **Briefing:** Presentation of the goals for the exploratory usability test. When applicable, we presented an overview of the main findings and decisions from the previous meeting.
- **Exhibition of visualisations:** Presentation of the visualisation prototypes and their features.
- **Free exploration:** Experts were invited to explore the visualisation prototypes and interact with their functionalities.
- **Debriefing:** Final considerations and review of requirements, ideas and potential suggestions.

It was important to not introduce any comments that could bias the results of the exploratory usability tests. It is acknowledged that this measure leads to more valid test results, as well as interesting comments made by users (Nielsen 1993).

Chapter 4 Analysis of datasets

4.1 Exploration of datasets

We took advantage of three datasets from different sources, which contain data from travel planning (retrieved from Move-me); bus ticketing and traffic counter sensors for the city of Porto. The following subsections describe the context of each dataset and the results of the exploration as defined in Section 3.1.1.

4.1.1 The Move-me dataset

Move-me is a mobile application developed in 2011 by OPT S.A that intends to improve the quality and access of information about public transport. The application requires an Internet connection, since it is capable of providing the location of the companies' vehicle fleet in real time. All users' requests made through Move-me are anonymous and stored into a database. Given that the application is a mean of providing information about transport, it is not possible to know if the user who made an information request actually travelled on the transport network. Move-me currently offers three types of public transport information:

1) Next departures: *provides the next bus/train/metro departures for a specific stop selected by the user.*
2) Route finder: *for a given origin and destination selected by the user, it provides a sequence of bus/train/metro routes that should be taken in order to reach the destination.*
3) Near stops: *provides the nearby stops based on the location provided by the user (assisted by the user's mobile phone GPS).*

Figure 17 shows three screens of the Move-me interface: home menu, next departures for a specific stop and a route plan between two locations.

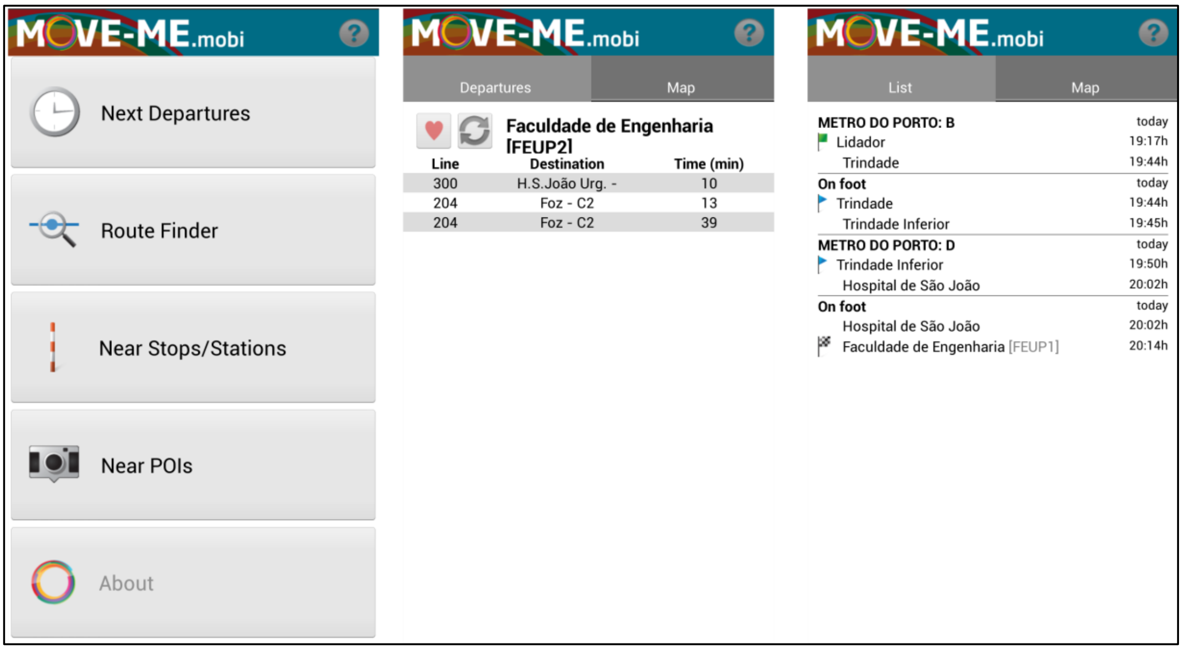


Figure 17 - Move-me mobile application

The dataset contains data from the period between January/2013 and December/2013. It consists of a single table from Move-me database that contains users' information requests, where each row was conceptually regarded as a *travel intention*, i.e. an intention to use public transport. In addition, it is assumed that the time of information request is the same as the desired travel time within the transport network. Every travel intention is tied to a *timestamp*. A row is also tied to the *type of request* (according to service that was requested).

Table 1 lists the relevant attributes of this dataset considered for this research. Therefore, a row is associated to either a request for next departures, or an arbitrary route plan or near stops. Each row contains a user's input that depends on the type of information.

For an information request for near stops, the input consists of two elements: a coordinate pair of geographic coordinates in WGS-84[5] format, which represents the current user's location or any other location selected by the user, and the desired search radius in meters. For example: (41.1306,-8.5867,500).

For an information request for next departures, the input is the name of a stop. For example: "Casa da Música". Finally, for a request for directions of a route plan, the input is a pair of origin and destination points, along with a set of intermediate passing points, if desired by the user. Given that this input is more complex in comparison with the other two input types, the request input is received in XML format, as depicted in the example below:


```xml
<?xml version="1.0" encoding="utf-16"?>
<Request>
  <StartTime>2013-11-
11T18:50:22.16+00:00</StartTime>
  <EndTime>2013-11-11T23:50:22.16Z</EndTime>
  <Type>RouteFinder</Type>
  <Route>
    <Track>
      <In>
        <Name>São Bento</Name>
        <Code>São Bento</Code>
        <Type>Stop</Type>
       <Provider>METRO DO PORTO</Provider>
      </In>
      <Out>
        <Name>Francelos</Name>
        <Code>CP_Francelos</Code>
        <Type>Stop</Type>
       <Provider>CP</Provider>
      </Out>
      <VisitTime>0</VisitTime>
      <ProviderCode />
    </Track>
  </Route>
<MaxResultTrips>3</MaxResultTrips>
</Request>
```


Table 1 - Structure of relevant attributes of the Move-me dataset

[5] World Geodetic System

Attribute	Type
Request date/time	Timestamp
Type of request	Categorical: • Next departures • Route Finder • Near stops
User's request input	Depends on the "Type of request": "Next departures": - Stop name "Route Finder": - Origin - Destination - Intermediate passing points "Near stops": - Latitude - Longitude - Radius
Response to request	Depends on the "Type of request": "Next departures": - List of next departures and remaining minutes "Route Finder": - Route plan including public transport services to be taken from origin to destination "Near stations": - List of nearby stations

Data primitives from the "Near stops" service already depict the spatial dimension as geographic coordinates. However, primitives from the other two services required additional manipulation (geocoding), given that a stop name itself represents a given location only at the conceptual level. We used an auxiliary dataset also provided by Move-me from 2013 that contained the geographic coordinates for each stop name.

4.1.2 The STCP dataset

The electronic bus ticketing system used by STCP consists of contactless travel cards named Andante and validating machines installed on all buses. Those Andante cards consist of two types: occasional and monthly card. Passengers with occasional cards pay a specific price for each travel. An occasional card can hold multiple travel tickets within it. Passengers with monthly cards pay a monthly fee and are entitled for travelling across specific city zones, which are specified during the moment of purchase. Also, Andante cards are multimodal, which means that they are valid for travelling across the Metro, train and tram networks.

An Andante card has a unique serial number. Whenever a passenger boards a bus, he/she must validate its ticket by moving it closer to the validating machine (see Figure 18). It is not needed to validate the ticket at the end of the journey.

The dataset collected from STCP concerns the months of January, April and May 2010. The dataset consists of a single table containing bus-ticketing data for all bus lines in Porto. Each row is regarded as a *travel event*, i.e. the realization of a travel into the public transport network on a specific line, in which a passenger places his/her ticket onto the machine. Every travel event is tied to two *timestamps*: the first indicates the time when the bus started the trip (from the terminus). The second indicates the time when the passenger validated his/her

Andante card. A row is also tied to a *vehicle number*, *line number*, *line direction* and naturally the *stop* where the validation occurred. Table 2 lists the relevant attributes of this dataset.

Figure 18 - Validation of Andante Card (Source: http://www.stcp.pt)

Table 2 - Structure of relevant attributes of the STCP dataset

Attribute	Type
Andante Card ID	Numeric
Andante Card Type	Numeric
Start Date/Time of bus trip	Timestamp
Date/Time of ticket validation	Timestamp
Vehicle number	Numeric
Line number	Numeric
Line destination	Numeric (1 or 2)
Stop ID	Numeric

A pair containing *date/time of ticket validation* and *vehicle number* can uniquely identify a *travel event* within the transport network. A conceptual example for this dataset is the following: [1155-0338-0901;03/01/2010 03:00:03;03/01/2010 03:04:33;3491;3M;1;1490], which is read as *"A validation of an Andante Card occurred at 03:04:33 on the vehicle 3491 in service on line 3M towards Aeroporto at stop AAL2-Aliados, whose trip started at 03/01/2010 03:00:03"*. Nonetheless, in order to achieve thorough understanding of this example, two attributes required prior adaptation. First, a row *per se* does not indicate the line destination explicitly, given that it is coded as either *1 (outbound)* or *2 (inbound).* Hence, this posed the need of translating it into a textual indication of the line destination, which in this case is *Aeroporto*. The same fact applies to the *stop ID* (1490), which had to be assigned to a name and geographic coordinates.

This assignment took advantage of an auxiliary dataset containing a list of STCP stops in 2008. Considering the age of this dataset, there were issues such as empty fields, duplicate entries, stops that existed in 2008 but were decommissioned in 2010, or stops that had their name changed (e.g. TRD1 to TRD5). Those cases implied manual interventions in order to assign all stop IDs to their names accordingly. After the interventions, it was possible to use the auxiliary dataset from Move-me that contained geographical coordinates for all stops names. Again, the same interventions were needed, as the dataset is from 2013.

4.1.3 Traffic count sensors data

The Porto City Hall installed traffic counter sensors in streets in order to keep track of the number of vehicles for further analysis and studies. Essentially, a sensor is made of coils that are installed under street lanes. Whenever a vehicle passes over a coil, the resulting magnetic induction triggers an electric signal which increments the count of vehicles. Every 5 minutes, the sensor sends the count to a central computer. At the end of the day, the software responsible for receiving the sensor counts gathers the data into a plain text file. Figure 19 depicts an example of a traffic counter sensor installed on a street.

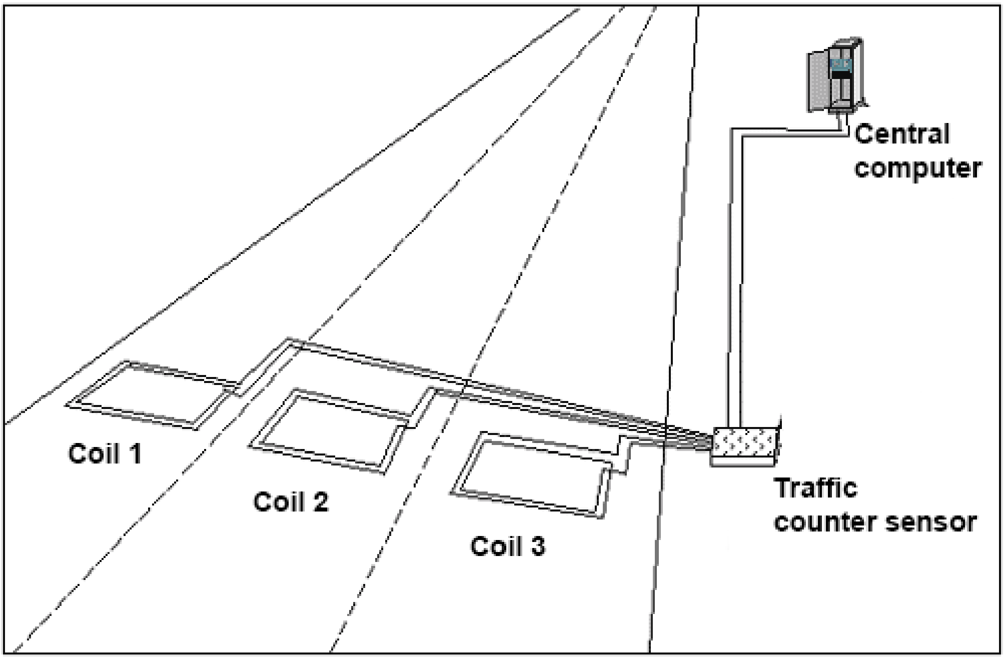


Figure 19 - Example of traffic counter sensor on a street. In this example, the sensor is linked to 3 coils, each at a street lane

The dataset collected from traffic counter sensors concerns the period between 2010 and 2013. Each set of streets belongs to a particular city zone, which are previously determined by city authorities. The dataset consists of over 15000 plain-text files containing daily counts for each sensor. The filename indicates the date and the respective city zone. For example, a file named 03012011-05.txt contains traffic flow data from January 3rd, 2011 for city zone 5. Each file line (row) is tied to a *time of the day* and *vehicle count* for each of the sensors installed on that zone. Table 6 lists the relevant attributes of this dataset.

Table 3 - Structure of relevant attributes of the vehicle traffic dataset

Attribute	Type
Date*	Date
City zone*	Numeric
Zone name*	Text
Time	Time
Counts for Sensors 1…N	Numeric
* Attributes do not belong directly to file’s contents and required manual adaptation	

Table 4 - Conceptual example for the traffic flow dataset

	Time	Sensor 1	Sensor 2	…	Sensor N
January 3rd, 2011 Zone 5	15:00	15	19	…	0
	15:05	11	5	…	3
	15:10	13	9	…	2

Given that some of the attributes did not belong to the files’ contents and there were many text files involved, it would be unfeasible to query data from specific time intervals. So as to overcome this issue, we developed a parser in order to port all raw data to the RDBMS. Essentially, this parser performed a loop during all the days of the considered years. For each day, the respective file was opened. Then, day and times were merged in order to form timestamps that would be associated to each of the readings from the sensors.

Another issue was the lack of geocoding for sensor locations. A workaround consisted in creating an auxiliary dataset, which contained geographic coordinates for each sensor. That was possible by taking advantage of city planning information including maps and diagrams retrieved from city authorities. Based on that information, we manually geocoded the locations for each sensor and inserted them into the auxiliary dataset.

4.2 Data characterisation

The pyramids depicted in Figure 20, Figure 21 and Figure 22 formalise the conclusion drew from the exploration of datasets.

Move-me dataset (Figure 20**)**

Theme: Information request about the public transport network. Its contents vary according to the service requested by the user.

Location: Variable, depending on each service.

> *Near Stops service*: consists of all possible coordinates within Porto, thus suggesting a continuous spatial domain.
>
> *Next Departures service*: consists of a discrete spatial domain $[s_1, \ldots s_n]$ of all bus/metro/train stops within Porto.
>
> *Route Planner service*: consists of a pair of origin and destination points, which are labelled with an address or stop name. Again, this suggests a continuous location domain.

Time: The instant when the request was made. For simplicity, we assume that the request time and travel intention time are equal.

Object: Travel intention within the public transport network. An immediate interpretation can be summarised as: *an intention to use the nearest public transport available within an area* (Near Stops); *an intention to be informed about the next departures from a given location within the network* (Next departures); *an intention to know the different ways of traveling between two points of Porto* (Route plan).

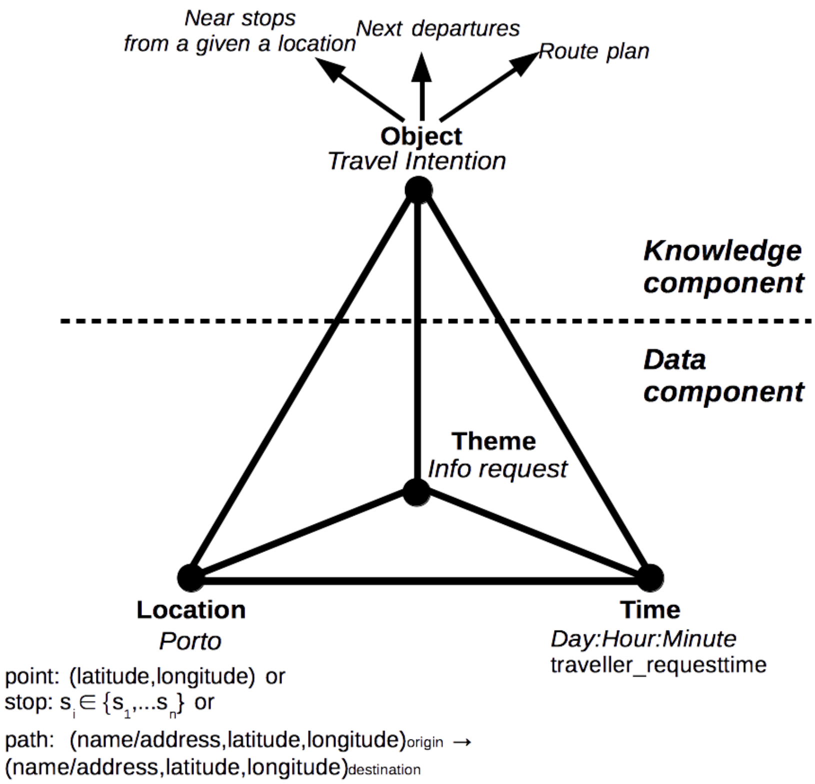


Figure 20 - Pyramid model for Move-me data

STCP dataset (Figure 21**)**

Theme: Ticket validations on the STCP bus route network.

Location: Consists of a discrete spatial domain $[s_1, \ldots s_n]$ of all STCP bus stops in Porto.

Time: The exploration in Section 4.1.2 implied that the time perspective can be branched into two components, given that each validation is tied to two time perspectives: the *time when the passenger validated the Andante card* and the *time in which the bus trip started (from the terminus point)*.

Object: If the first time perspective is considered, the object can be interpreted as a *single travel event within the transportation network*. The second time perspective allows interpreting an object as *all travel events that happened on a certain bus trip*.

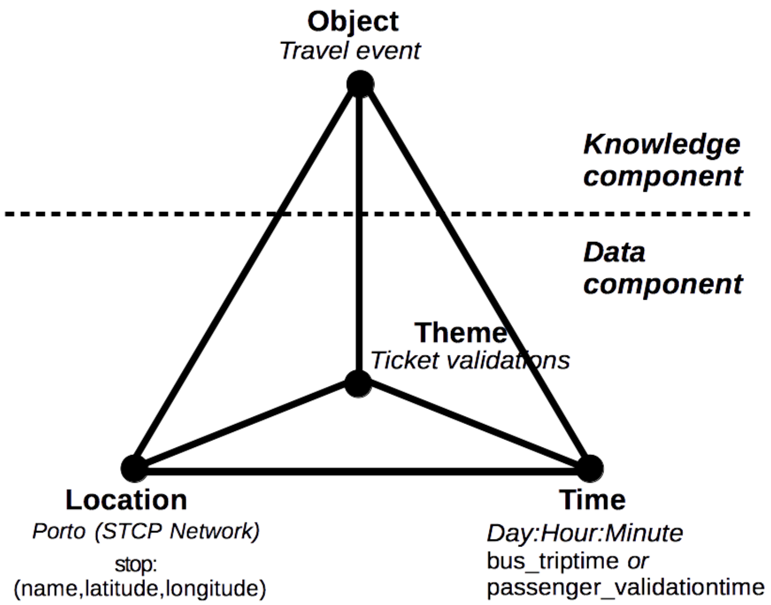


Figure 21 - Pyramid model for STCP data

Traffic counter sensors dataset (Figure 22**)**

Theme: Count of vehicles detected by a sensor.

Location: Consists of a discrete set $[l_1, \ldots l_n]$ of locations for each sensor. Each sensor belongs to a pre-defined city zone, which is labelled with a code and name.

Time: One of the possible 288 5-minute intervals for each day, which sum 24 hours.

Object: Traffic flow on a specific street during a 5-minute interval.

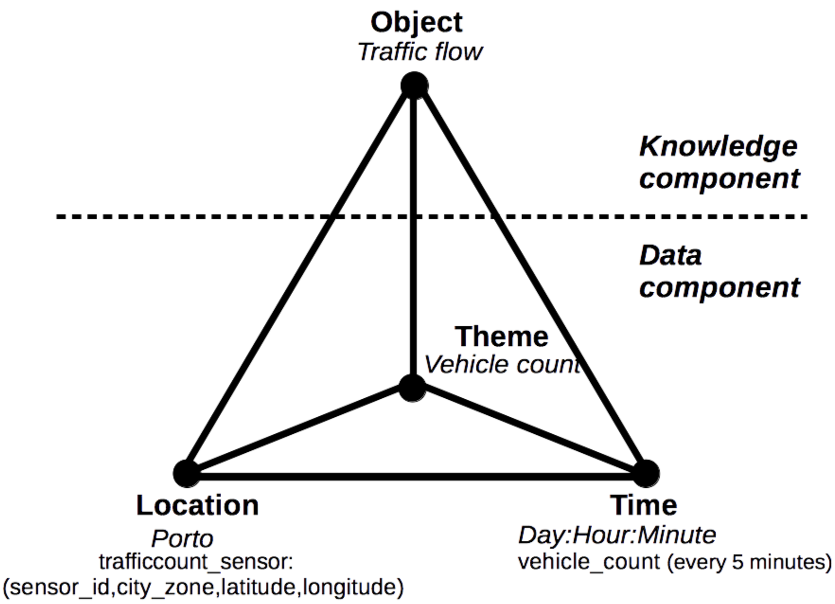


Figure 22 - Pyramid model for vehicle traffic data

The exploration of datasets and data characterisation allowed identifying that the data primitives from the three different datasets share structural characteristics in terms of *spatial domain* and *weight* (see Table 5). It follows that they could be compatible with the same visualisation techniques from a technological perspective. However, from a conceptual perspective (concretely from the experts' viewpoint), it would not possible to infer immediately if those visualisations would be adequate, thus naturally requiring further evaluation with experts. Possibly, a visualisation might not suit two different data types simultaneously if there are substantial differences at the conceptual (theme) level.

We define *spatial domain* as the set of possible locations that a data unit can be associated with. For example, a data unit from a Move-me request for near stops can be tied to any geographic location, as this information is directly retrieved through GPS. By establishing an analogy with mathematical numeric intervals, we can infer that the spatial domain for this data unit is essentially *continuous*, considering the precision of GPS systems.

On the other hand, for example, a data unit from a Move-me request for next departures is tied to a discrete set of stops that are previously determined by transport authorities, where each stop is tied to a unique geographic location. This allows us to infer that the spatial domain for this data unity is *discrete*.

Finally, a data unit can be associated with a spatial domain consisting of origin and destination points, as well as intermediate passing points, which resemble the concept of a graph. This naturally implies the concept of direction. We classify this type of spatial domain as a *directed graph*.

We define *weight* as the magnitude (a scalar quantity) of a geographic location, i.e. a stop or passengers' location. For example, the concept of weight can be used to suggest that locations with higher magnitude have increased demand.

Table 5 - Classification of structural similarities between data primitives

Spatial domain	Discrete
	Continuous
	Directed graph
Weight	Weighted
	Non-weighted

Chapter 5 Results of the User-Centred Design Process

This chapter describes the results of the User-Centred Design process. The prototypes presented herein depict specific examples that draw attention to the distinctive characteristics of each visualisation. After developing an initial set of visualisation prototypes, we approached experts in order to understand the potential application of visualisations within their context of use. We held exploratory usability tests with experts during every meeting in order to gather their inputs, so they could be considered during further iterations of the prototyping phase. The iterations of the conceptual development phase resulted on the analysis of 16 hours of audio recordings. High-resolution images for some visualisation prototypes are found in APPENDIX C.

5.1 Context of Use Analysis

Experts stated that Information Visualisation would benefit them in the following activities:

- Analysis of trends in urban mobility throughout time (e.g. detecting lines or stops that could be decommissioned/created; detecting the increase/decay in traffic flow on specific locations throughout time)
- Social sensing and analysing ridership levels (e.g. passenger/vehicle traffic in certain areas, lines and stops)
- Analysis of specific areas or stops (e.g. detecting areas with higher concentration of potential passengers and detection of extreme and seasonal events)
- Collaboration with urban or market researches (e.g. identifying areas or lines/stops that may be more valuable for advertising purposes or traffic signalling)
- Support decisions in traffic/route deviations

Experts also stated that visualisations would be useful for monitoring and retrieving *snapshots* of the current condition of the transportation network/traffic on a specific day and time. We also considered technological factors, given that visualisation proposals should be implemented on a future phase. This reflected the choice for prototyping with free tools that could communicate with databases successfully and render visualisations of large datasets adequately. The same tools are eligible to use on a future implementation.

5.2 Prototyping for Domain Visualisation

According to Section 3.2.2, we developed vertical prototypes for the iterations of the Domain Visualisation phase. The results from the analysis of datasets described in Chapter 4 allowed identifying structural characteristics shared between data primitives, thus allowing particular visualisations to support multiple data types. Table 6 lists the 11 visualisations developed for this study and the structural data characteristics they support. For each visualisation, we also present a formal time modelling as described in Section 3.1.3, along with its four perspectives, which allowed us to determine how time should be depicted within each visualisation.

Table 6 - List of visualisations developed for this study and the data they support

Visualisation	Structural data characteristics				
	Spatial domain			Weight	
	Discrete	Continuous	Directed graph	Weighted	Non-weighted
Circle markers	●				●
Clustered markers		●			●
Heat map	●	●		●	●
Sized circles	●			●	
Calendar heat chart	●			●	
Line plots	●	●		●	●
Radial heat chart	●			●	
3D Lines (*experimental*)			●	●	●
Radial convergence layout			●	●	
Geographic Graph			●	●	●
Coupled Visualisation	●			●	

Experts visualised data from different periods by using a visual query control (see Figure 23). It also indicates the period and the number of data units (rows) that were being visualised at that moment. The control only accepts date and time inputs according to the formats specified in the same figure, preventing input errors.

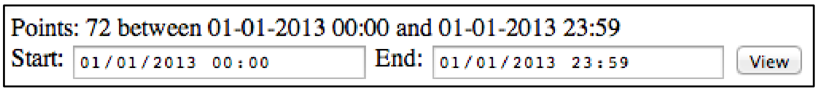


Figure 23 – Visual query control

Visualisation 1: Circle markers

Scale: discrete	**Scope:** point-based	**Arrangement:** linear	**Viewpoint:** ordered

This visualisation consists in representing each data unit as a circle. In particular, Figure 24 shows an example of this visualisation coupled with data from Near Stops service, which has a continuous spatial domain. Each circle represents a request for nearby stops made by a traveller for a particular location. Areas with higher density of circles suggest higher flows of potential passengers looking for a stop. The same figure shows the difference in passenger flows in the surroundings of Matosinhos in July (left) and November (right) 2013 respectively.

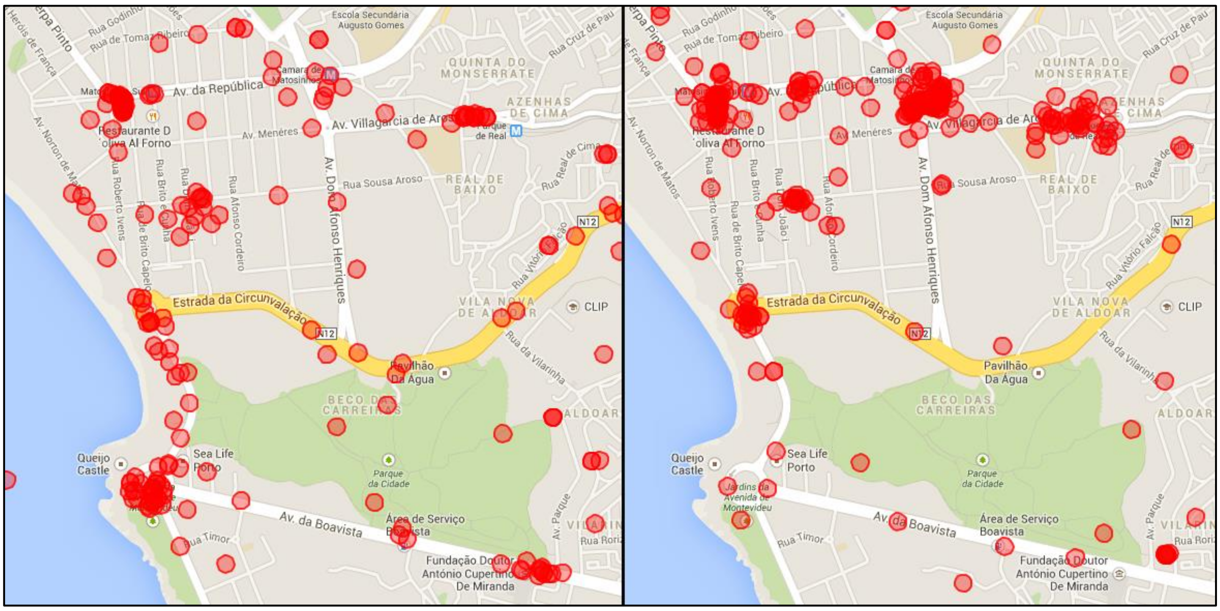


Figure 24 - Circle Markers for July 2013 (left) and November 2013 (right)

On summer, potential passengers concentrate at areas such as Queijo Castle and Av. Norton de Matos, a coastal avenue. In November, there is a higher concentration along Av. da República, especially near Metro stations. This area is known for its restaurants.

Visualisation 2: Clustered Markers

Scale: discrete	**Scope:** point-based	**Arrangement:** linear	**Viewpoint:** ordered

This visualisation consists of grouping data units into clusters according to the zoom level that is applied. The goal was to reduce visual clutter and to provide a colour coding based on a discrete and ascending scale, consisting of blue, yellow, red and purple (see Figure 25). It is possible to adjust the range of numeric intervals that are assigned to each colour.

When zooming into an area, clusters unfold into smaller clusters. Eventually, a single marker (denoted as a red passenger icon) may not belong to a cluster if it is not sufficiently close to a cluster. Figure 26 shows the same location and period depicted in Circle Markers.

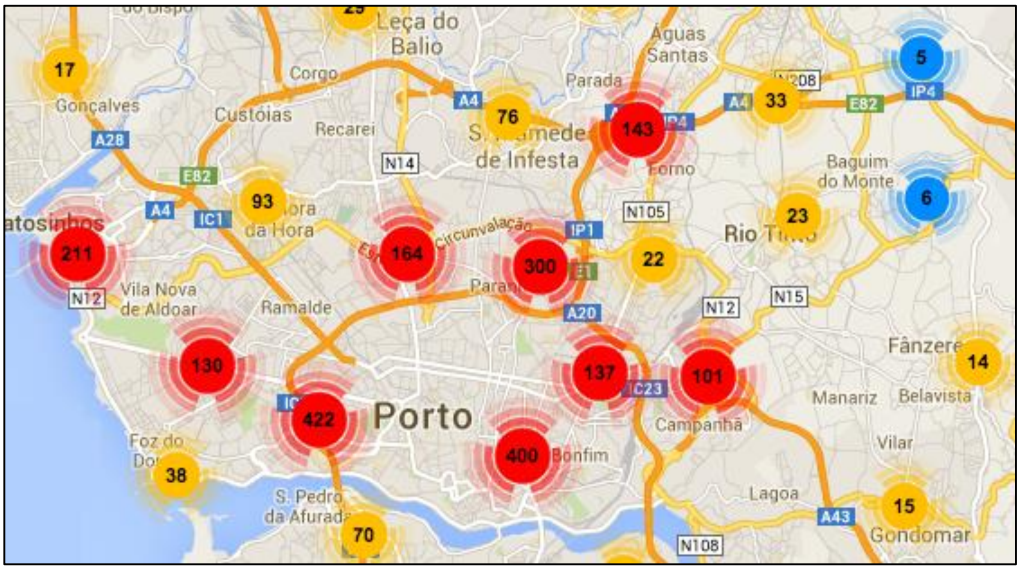


Figure 25 - Overview of Clustered Markers visualisation

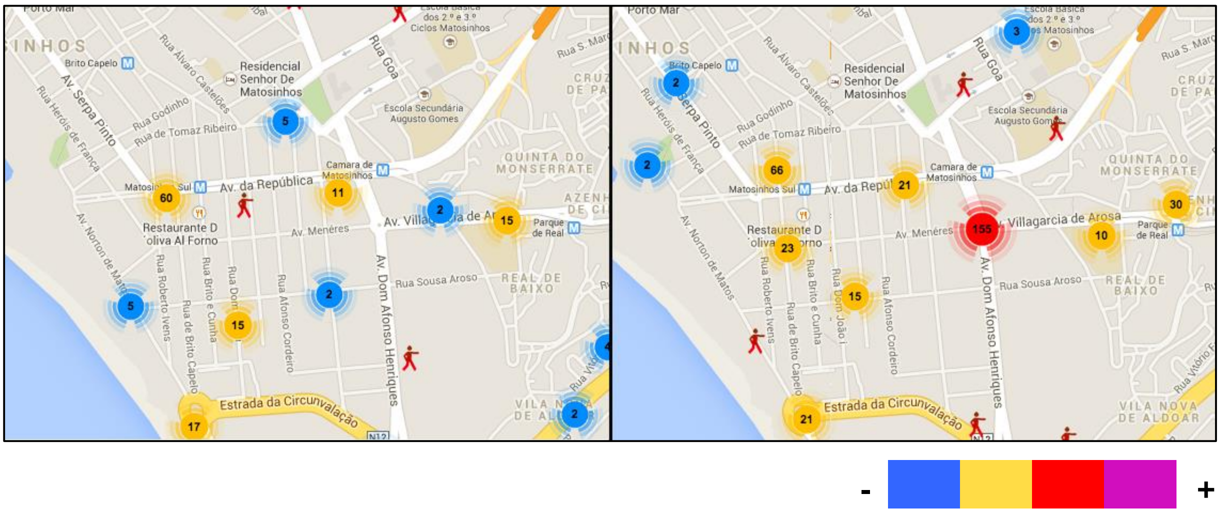


Figure 26 - Clustered Markers for July 2013 (left) and November 2013 (right)

Visualisation 3: Heat map

Scale: discrete	**Scope:** point-based	**Arrangement:** linear	**Viewpoint:** ordered

Heat maps are widespread on scientific visualisations where an object property (e.g. temperature) is colour-coded, and then projected onto the object itself (see Figure 7 right). Here, the concept of "heat" suggests areas with more user requests for nearby stops. Again, Figure 27 depicts the same location and period.

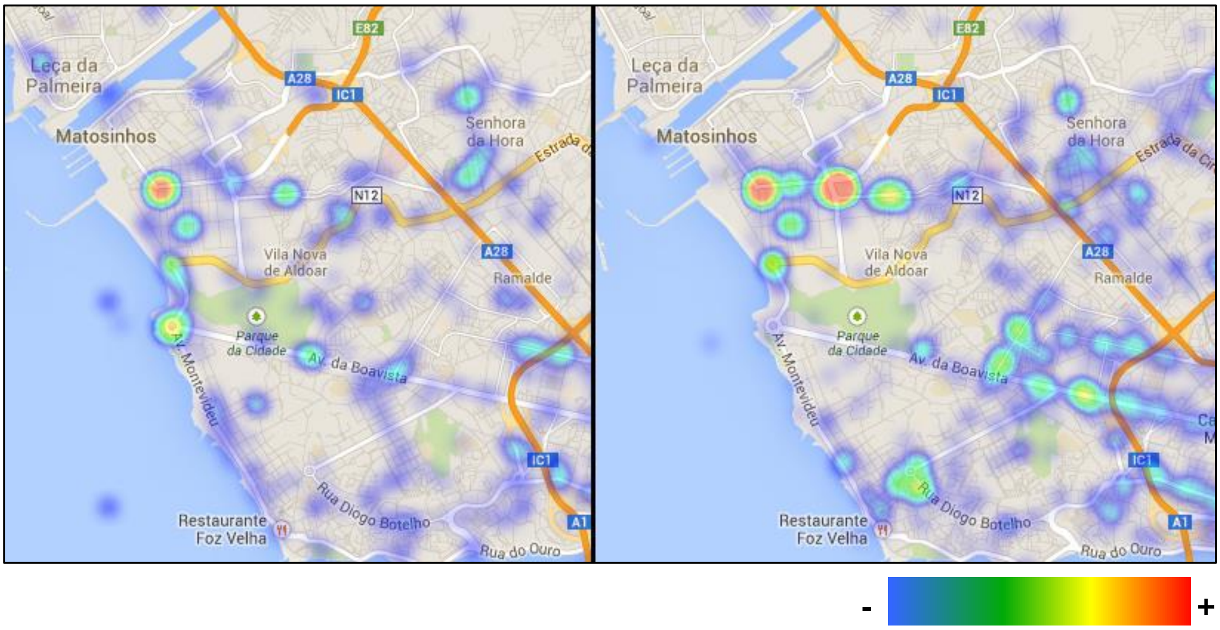


Figure 27 - Heat maps for July 2013 (left) and November 2013 (right)

The same visualisation technique can be used to represent data with discrete spatial domains. Hence, each point represented on the visualisation is associated to a weight defined by the number of reoccurrences of that point. Figure 29 and Figure 30 depict data from the Next Schedules service for St. John's Night and for vehicle flows at Bolhão city zone.

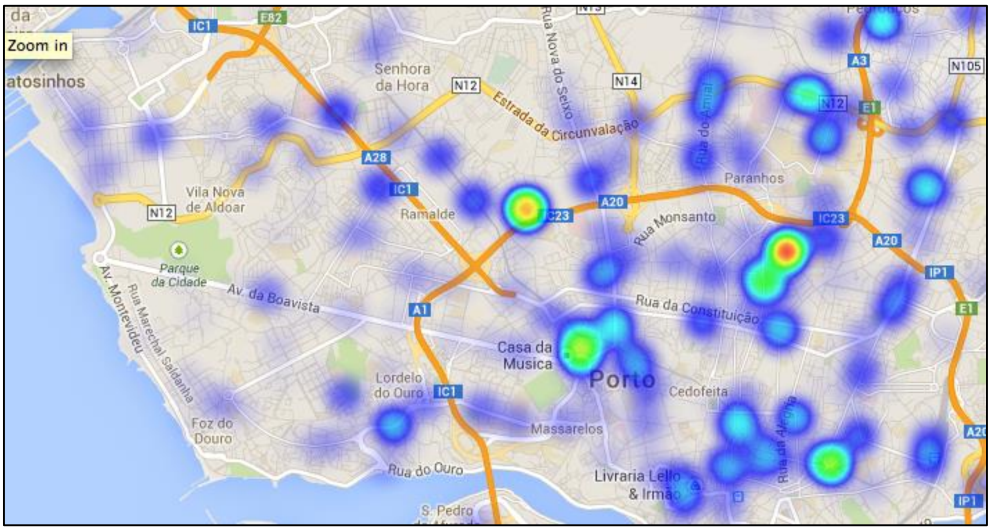


Figure 28 - Overview of a heat map representing schedule requests for a given day in Oporto

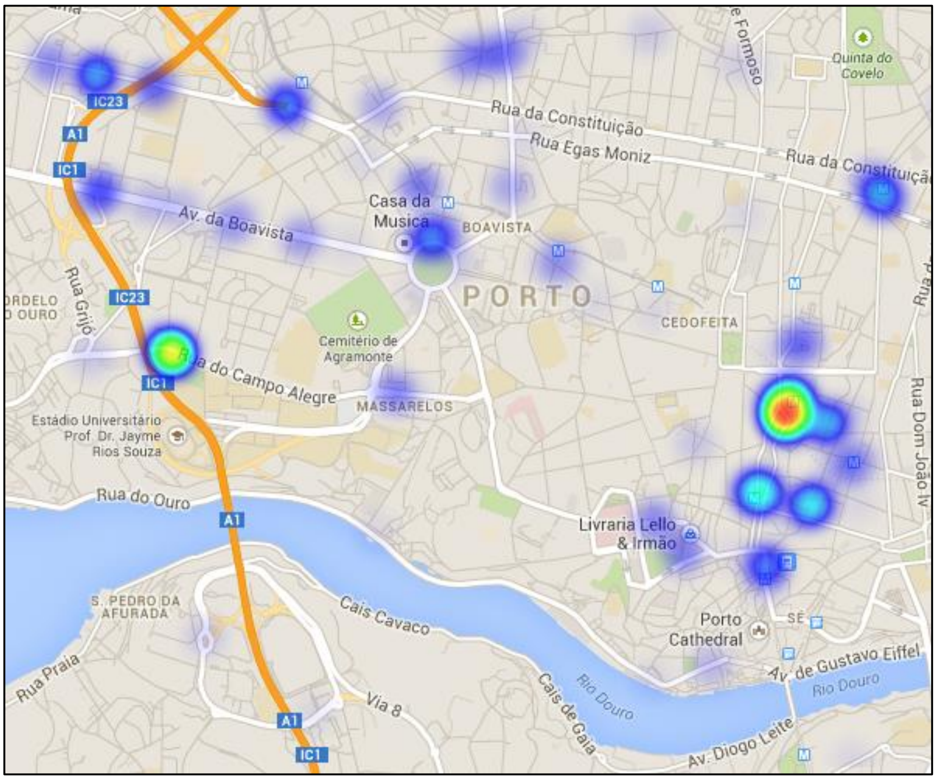


Figure 29 - Heat map for St. John’s Night 2013

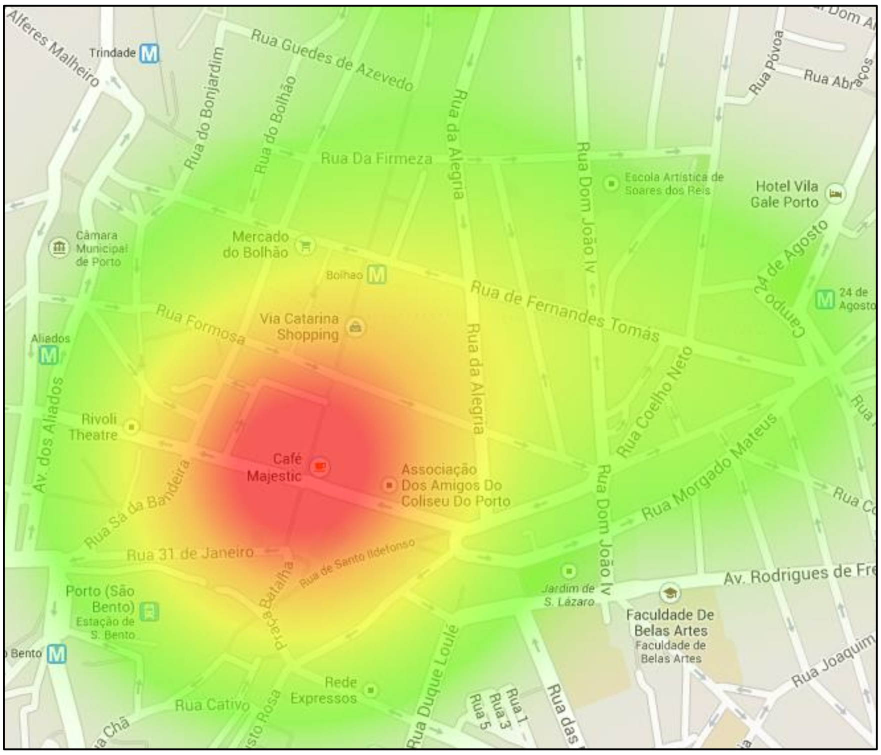


Figure 30 - Heat map for traffic flows at Bolhão city zone for a specific day

The preliminary colour-coding strategy consisted on normalising magnitudes by mapping them to the [0,1] interval. After, the normalised values were linearly interpolated to the RGB[6] colour space (see Figure 31).

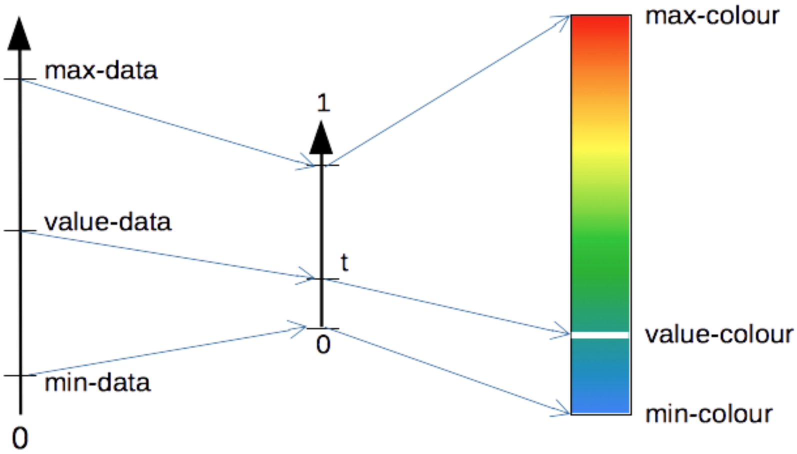


Figure 31 - Preliminary colour-coding strategy for heat maps

[6] Red-Green-Blue

Visualisation 4: Sized circles

Scale: discrete	**Scope:** interval-based	**Arrangement:** linear	**Viewpoint:** ordered

This visualisation is a variant of Circle markers, where each data unit is represented as a circle with a radius that depends on a certain parameter, here referred as magnitude. In this example, the coordinates of a stop determine a circle's centre. The magnitude is the number of 'next departures' requests made by users. Hence, larger circles suggest stops with higher demand. Circles are colour-coded and represent different transport providers. Yellow and blue represent Metro do Porto and STCP stops respectively. Red circles represent stops belonging to other transport operators. Figure 32 depicts an example of Porto during Valentine's Day (between 6pm and 12am) and St. John's Night (between 6pm and 5am).

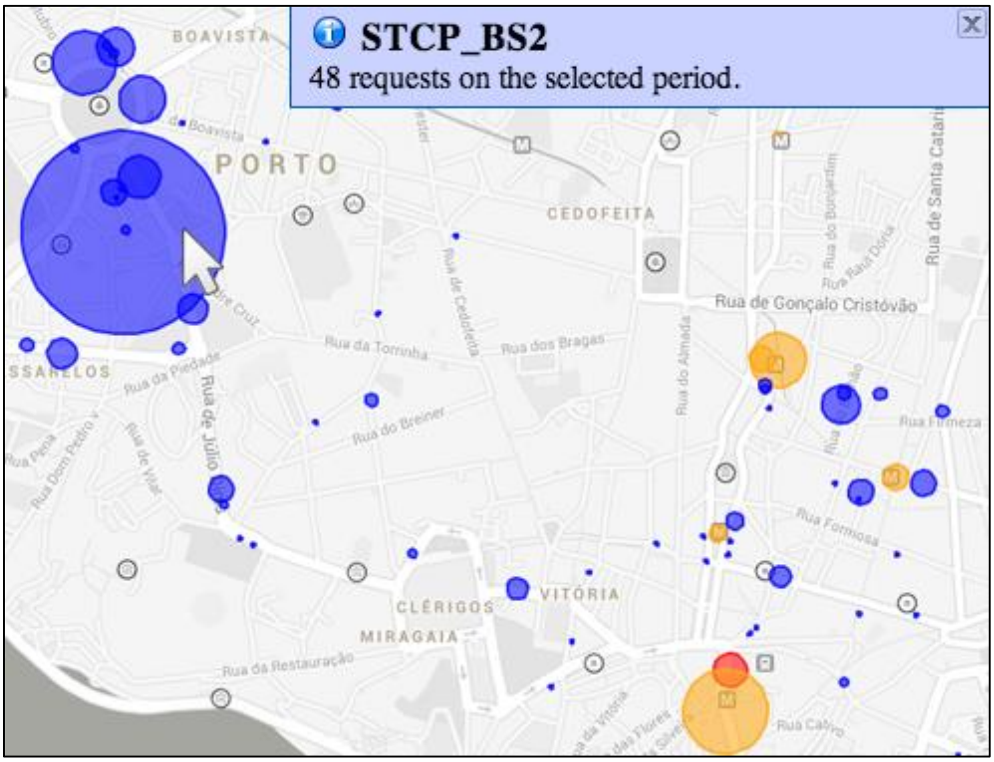


Figure 32 - Density circles for Valentine's Day 2013

There is a high volume of requests for next schedules near Boavista area (top-left). The most requested stop is STCP_BS2 with 48 schedule requests. This visualisation is interactive: by clicking on a circle, a dialog indicates the number of requests that were made during the query period (see Figure 32).

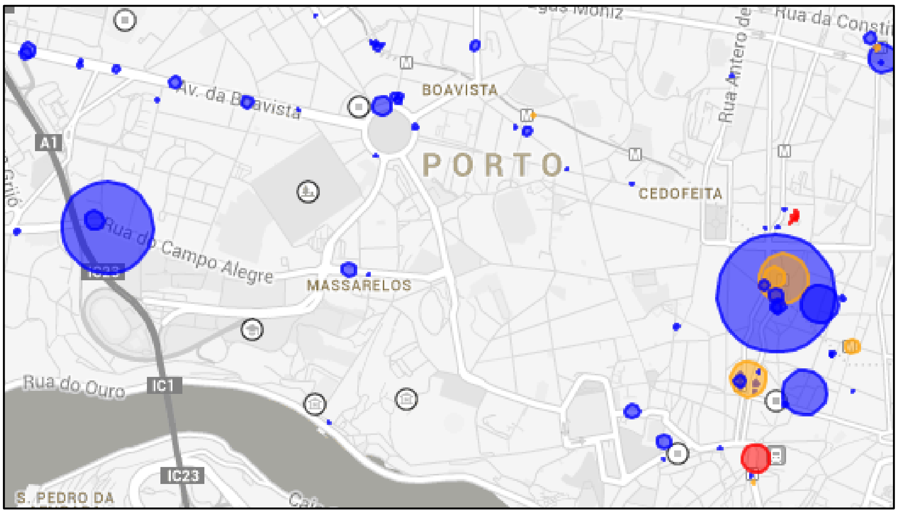


Figure 33 - Density circles for St. John's Night 2013

During St. John's Night, there were concentrations around the stops in downtown (bottom-right), where the festival took place. Another concentration is evident at Campo Alegre area.

Visualisation 5: Calendar heat chart

Scale: discrete	**Scope:** interval-based	**Arrangement:** linear	**Viewpoint:** ordered

This visualisation aggregates data primitives according to the day they are related to. Months are aligned horizontally, and each month column corresponds to a day of the week, from Monday to Sunday. Magnitudes are linearly interpolated by the same method described in Figure 31. In Figure 34, each event is a request for next departures made for a specific stop. Placing the cursor over a specific day shows the total number of requests for that day. Brighter colour tones indicate higher number of requests. Extreme events (outliers) can reach white colours. Figure 34 exemplifies this fact. Due to a national strike on June 27th 2013, which included the transport network, the number of requests for next departures for Trindade metro stop reached extreme levels in comparison to the other days.

Figure 34 - Calendar heat map for Trindade Metro stop with focus on a specific date

Visualisation 6: Line plots

Scale: continuous	**Scope:** interval-based	**Arrangement:** linear	**Viewpoint:** multi-perspective

In Figure 35, each line represents the evolution of information requests for next departures made through Move-me for each month and specific stop. It is possible to compare multiple stops in the same visualisation. Line styles consisted of smooth Bézier curves.

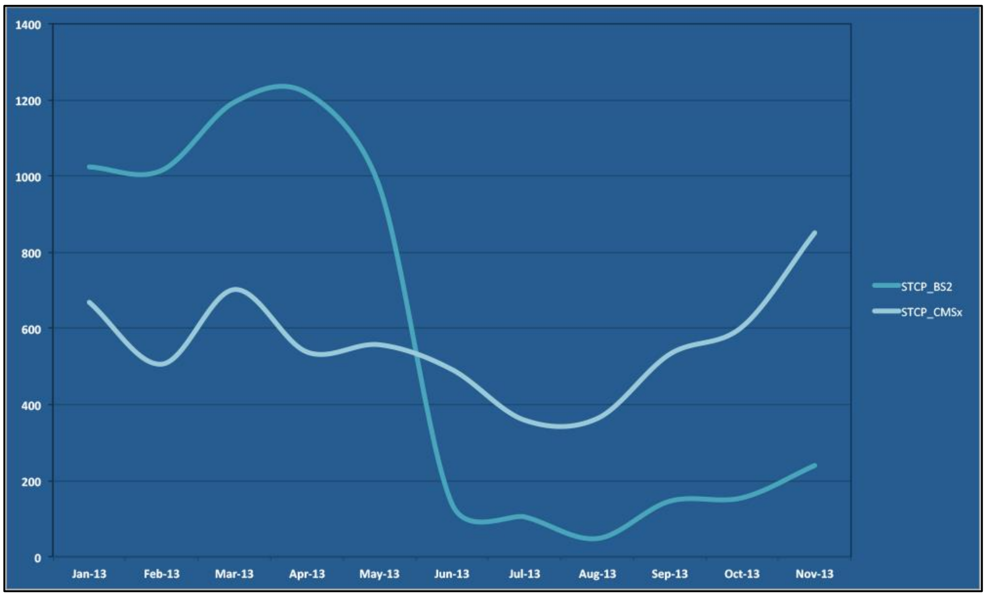


Figure 35 - Line chart comparing schedule requests between bus stops

Line plots were also coupled with ticketing data to represent the evolution of average number of validations per stop over a certain period, for a specific bus line. In Figure 36, bus stops are depicted on the horizontal axis, along with the respective number of ticket validations on the vertical axis.

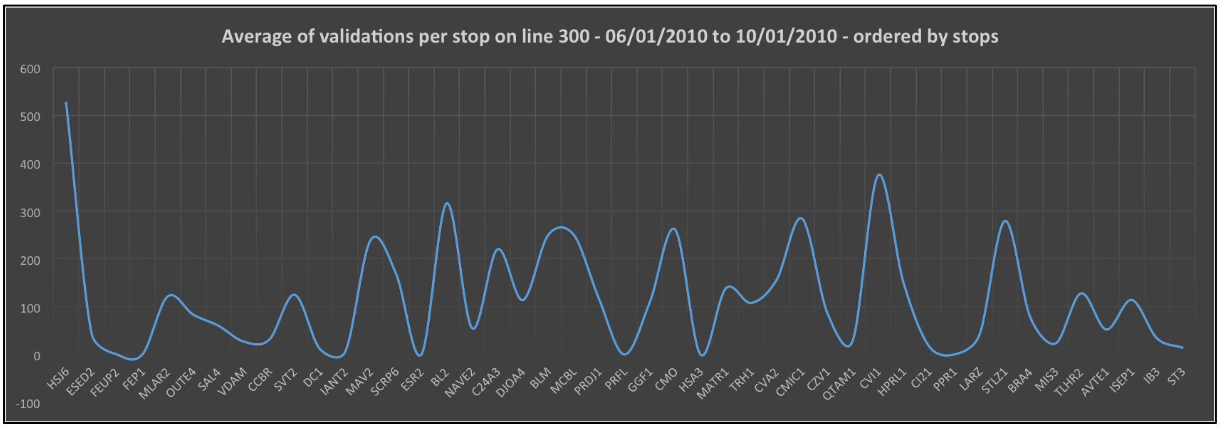


Figure 36 - Line chart of average validations per stop for a specific bus line

Visualisation 7: Radial heat chart

Scale: continuous	**Scope:** interval-based	**Arrangement:** cyclic	**Viewpoint:** multi-perspective

This visualisation aggregates data on a cyclic pattern, such as ticket validations and travel intentions from a specific stop. Radial layouts might better support the task of detecting patterns. The natural reading order (left to right) might cause false interpretations when using visualisations that follow linear layouts. Users tend to minimise eye movements; if graphic artefacts are spatially close, the sampling cost should be reduced (Keim et al. 2006, Ware 2004).

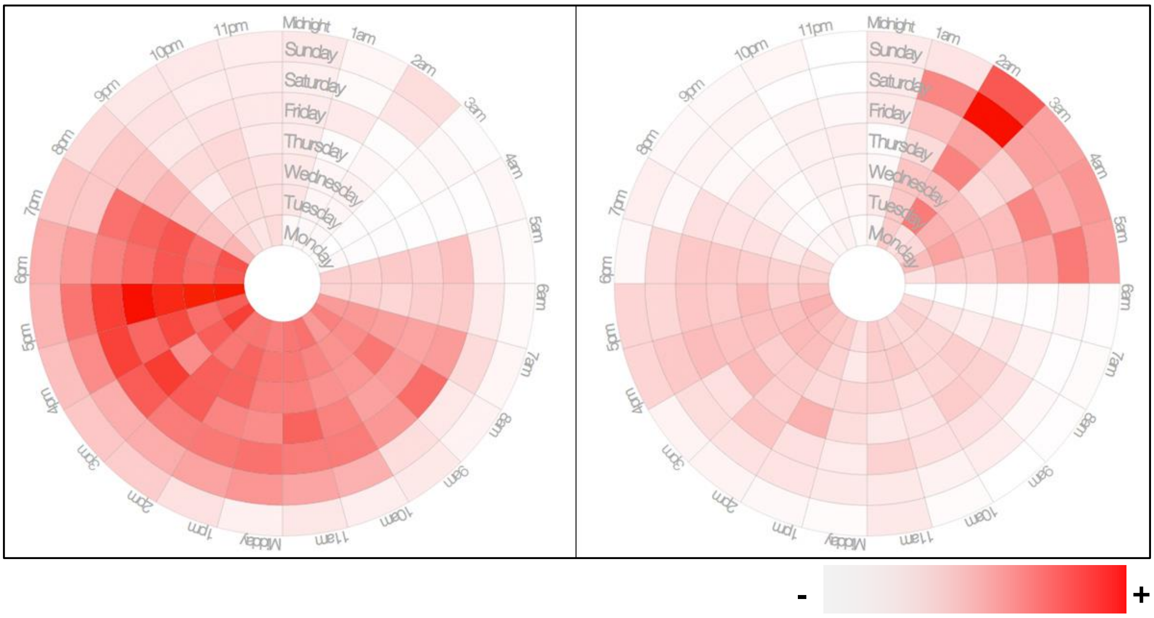


Figure 37 - Radial heat chart of ticketing data for two bus stops

Figure 37 represents ticketing data for two bus stops. Each circular row represents a day of the week, and each circular column represents a day hour. It is noticeable that the station on the left has higher demand during work hours, whereas the one on the right has higher demand during late night hours, especially on Saturdays between 1-2 a.m.

Visualisation 8: 3D Lines

Scale: discrete	**Scope:** point-based	**Arrangement:** linear	**Viewpoint:** ordered

This visualisation intends to represent data within a directed spatial domain. Each origin-destination pair is depicted as a straight line. In order to reduce the inevitable visual clutter brought by the nature of this visualisation, each line has a low opacity level. Hence, lines are depicted with high opacity when there is a significant overlap between them, which suggests a higher passenger flow on that track or points. 3D representations of data divide researchers, as they suggest that unnecessary difficulties such as information occlusion or hidden faces might appear (Aigner et al. 2011). Considering the aforementioned visual clutter, we considered adequate to introduce an additional dimension in order to facilitate the task of navigating through data. Figure 38 shows route planner requests in Porto on November 2013.

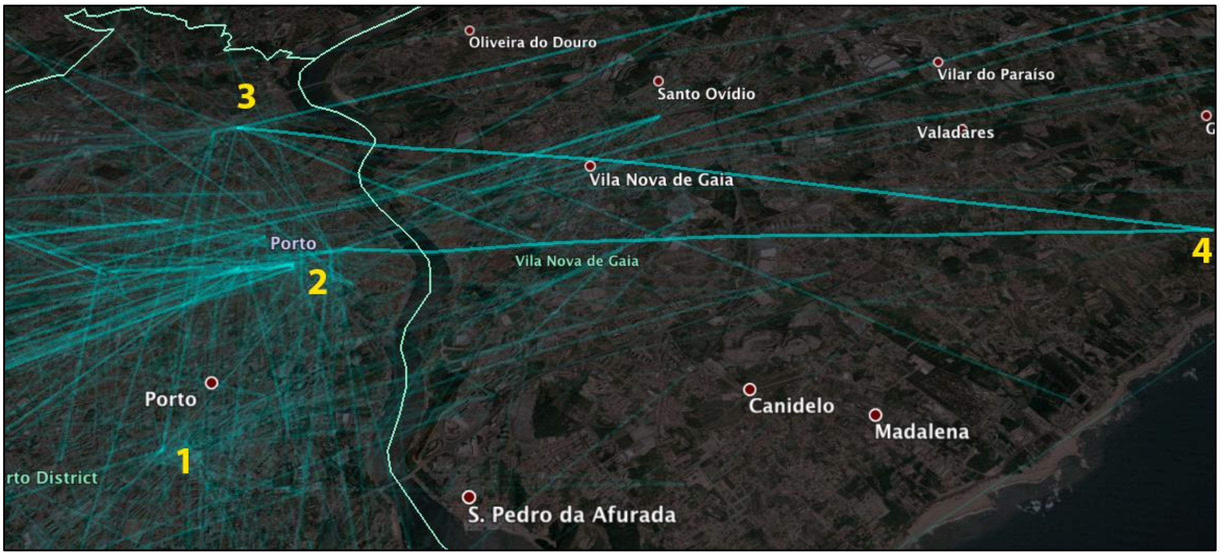


Figure 38 - Google Earth based visualisation

It is noticeable that there is a significant amount of route plans that include Casa da Música (1), São Bento (2), Campanhã (3) or Francelos (4

Visualisation 9: Radial convergence layout

Scale: continuous	**Scope:** interval-based	**Arrangement:** linear	**Viewpoint:** ordered

Circos[7] originally proposed this visualisation for representing genomic data, although it can be used for visualising relations between other data types, as in our case. This visualisation depicts travel intentions (route plans) between origin-destination pairs. Instead of representing stops according to their geographic location, they are organised into circular arcs proportional to the number of requests that involve them.

Chords, the coloured strips, represent travel intentions between two stops, and their weight vary according to the number of requests that are made through the route planner application, as in the Geographic Graph visualisation. Colours are randomly assigned to stops, and a chord's colour is mapped to colour of the destination stop. Figure 39 shows all travel intentions in Metro do Porto network on November 2010. A filtering is applied in order to hide segments that did not have significant amount of travel intentions.

[7] Website: http://circos.ca/

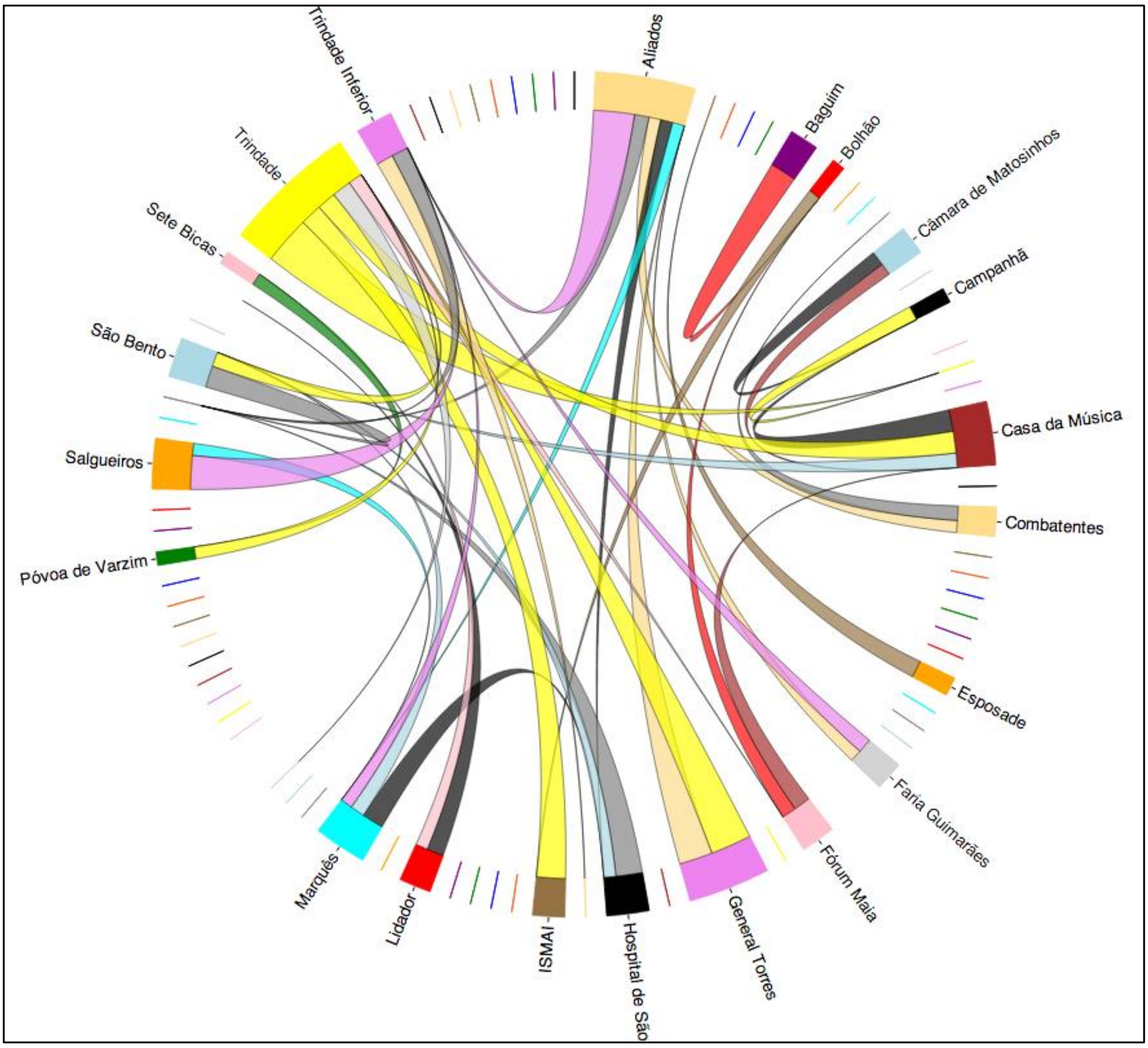


Figure 39 - Chord Diagram for travel intentions in Metro do Porto network (November 2013)

Chords inevitably overlap in this visualisation, which is usually regarded as 'hairball' in Information Visualisation. Hence, it was crucial to introduce an interactive way to reduce visual clutter and facilitate reading. By placing the cursor over a specific station (e.g. Trindade), the visualisation emphasises the chords related to that stop and occludes the remaining ones.

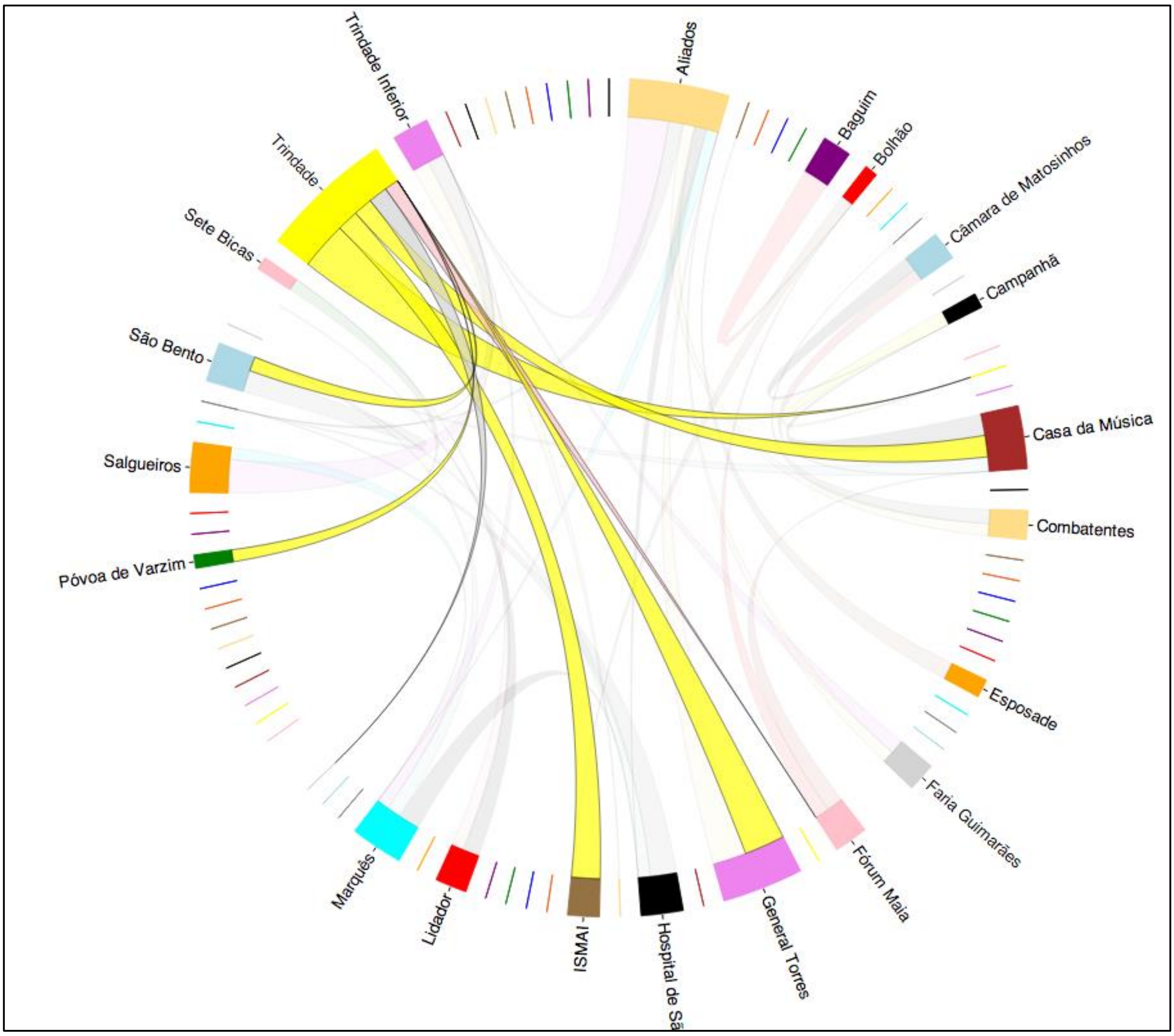


Figure 40 - Visual filter for the Chord Diagram (focus on Trindade stop)

Visualisation 10: Geographic Graph

Scale: continuous	**Scope:** interval-based	**Arrangement:** linear	**Viewpoint:** ordered

A transportation network can be understood as a set of nodes (stops) and edges (tracks). This visualisation intends to depict passenger flows across Metro do Porto network. Each line (edge) represents a set of travel intentions between two stops in both directions. The number of requests for that edge determines its line weight. Figure 41 shows an example for travel intentions on November 2010. In order to reduce visual clutter, a filtering process was applied, in which only edges above a certain weight would be visible. This causes some stops to be invisible in the visualisation. A clustering algorithm is responsible for colour coding. It does so by detecting clusters of similar travel patterns between stops. Each cluster is assigned to a random colour.

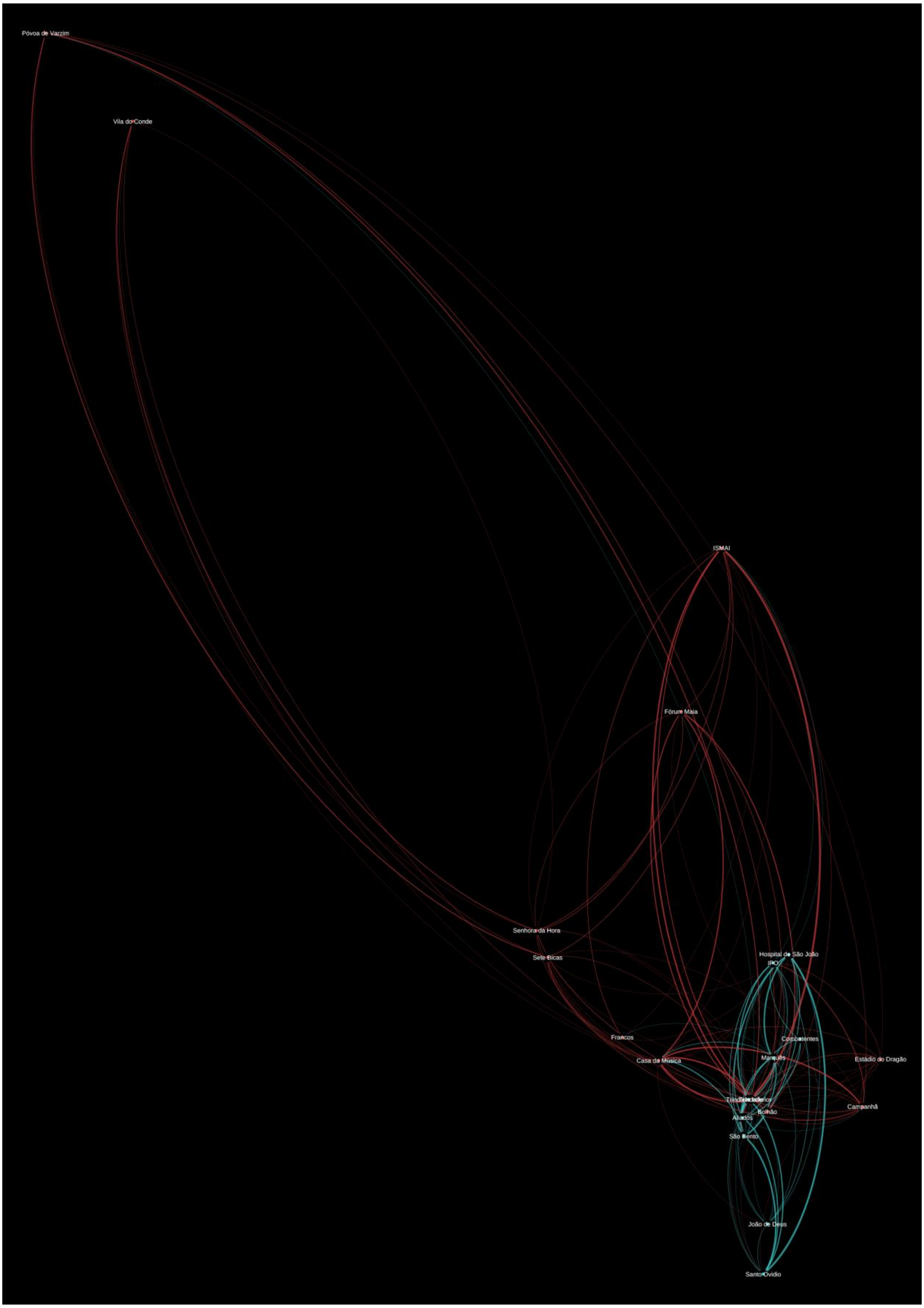


Figure 41 - Geographic Graph: travel intentions on the Metro do Porto Network (November 2013)

Visualisation 11: Coupled visualisations

After some iterations of the Domain Visualisation phase, experts stressed the importance of combining visualisations, especially when analysing historical data. Based on that statement and their request, we proposed an additional prototype for representing the historical evolution of the Annual Average Daily Traffic (AADT) for a specific set of traffic counter sensors installed in Porto, in the period between 2010 and 2013. This theme was chosen by some experts, who had special interest in analysing the evolution of traffic flows at those points.

The prototype coupled the Sized Circles and Line plot visualisations (see Figure 42). The visualisation starts by representing circles centred at each sensor's location. A circle's radius is defined by the absolute AADT variation in relation to the previous year, which can be represented as $r = |AADT_{year} - AADT_{year-1}|$. The absolute value naturally follows from the fact that a circle cannot have a negative radius. However, variation is depicted as colour. Positive AADT variations (increase in traffic flow in relation to the previous year) are represented in red. Negative variations are represented in blue.

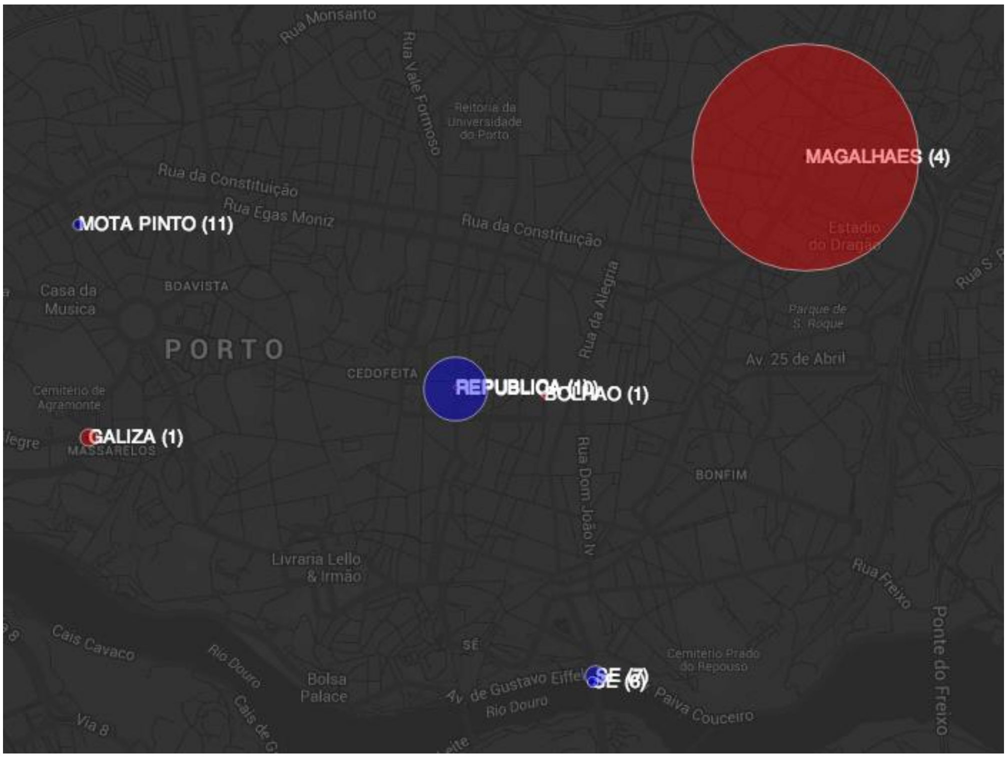


Figure 42 - AADT variation for traffic counter sensors between 2013 and 2012 (Sized circles)

The visualisation includes the same interaction features proposed in the other geographic visualisations such as zoom and pan. There is an intense increase in the number of vehicle flows detected by "Magalhães (4)" sensor in relation to the others.

Clicking on a circle allows a user to retrieve historical data for that sensor (see Figure 43). The historical data consists on line plots with markers indicating the AADT for each year.

When clicking on a marker for a specific year (e.g. 2011), the Sized circles visualisation automatically refreshes, showing AADT variation between 2011 and 2010.

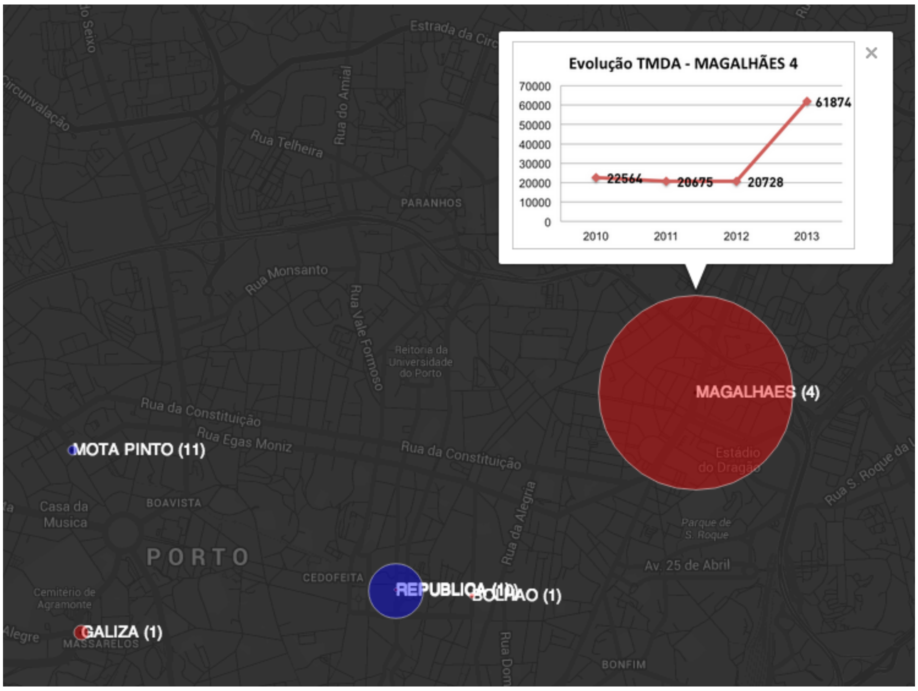


Figure 43 - Line plot for historical AADT data for a traffic counter sensor

5.3 Interaction and usability studies

The results from exploratory usability tests with experts were organised by visualisation and their impressions were classified according to the following facets:

- Readability: *effort required for understanding the concept behind the visualisation and its elements (e.g. symbols, layout). Also, it includes any impressions about the effort required for reading it and extracting knowledge.*
- Colour coding: *adequacy of the colour palettes used on graphic elements for representing data.*
- Visual clutter: *aspects concerning visual pollution in the visualisation.*
- Suitability: *(in)adequacy of the visualisation for representing the different types of data regarded in this research.*
- Requirements or requests: *user requirements or potential suggestions to be considered during further visualisation developments* (if applicable).

We also present the reworked prototypes based on experts' statements collected throughout the process, in case of significant changes applied to the visualisation concept.

<u>Visualisation 1 - Circle markers</u>

Readability: Some effort for understanding the visualisation concept, typically requiring additional explanations about it. Experts typically expressed negative reactions to it, stating that circles were potentially distractive.

Colour coding: The red colour for circles was considered inadequate and distractive. The map colour palette was considered acceptable, although monochromatic pallets would be preferable.

Visual clutter: Substantial indications of visual clutter, due to the use of a single graphic element and a distracting colour. In higher zoom levels, overlapped circles made difficult to identify street names and areas precisely. In lower zoom levels, the increased density of circles on the map led users to misinterpret data. For example, small concentration of circles (~10) was considered as dense as actual high concentration of circles (~80).

Suitability: Not suitable due to the severity of visual clutter. Experts stated that Circle markers fail to represent density, and it is unfeasible to use it in lower zoom levels. Moreover, they stated that the exact amount and location of people are usually irrelevant when identifying trends from data with a continuous spatial domain, thus this visualisation would have little or no applicability.

Visualisation 2 – Clustered markers

Readability: Some experts had initial difficulties to understand the visualisation concept, namely the automatic recalculation of clusters according to the zoom level. Still, experts were able to identify facts and draw conclusions about them.

Colour coding: The colour scale was considered inadequate by half of the experts. They stated that the blue colour resembles higher magnitudes than yellow. Moreover, they had difficulties to visualise yellow clusters due to the map colour palette. Experts required that the heat colour scale should be the same as in heat maps visualisation, and the map colour palette should be monochromatic.

Visual clutter: Low-magnitude clusters (blue) at lower zoom levels were considered distractive and irrelevant. Experts stated that it should be possible to filter the visibility of clusters according to an adjustable parameter such as *magnitude* itself.

Suitability: Clustered markers were considered adequate for representing and analysing trends at macro levels (city zones). The unfolding clusters at higher zoom levels turns unsuitable analysing trends in specific city zones. However, the current colour coding requires further improvement. Experts emphasised its significance for supporting decisions at the operational and strategic levels.

Requirements: The map colour palette should be monochromatic and incorporate dark tones.

The visualisation should allow filtering clusters according to a minimum size parameter.

The aforementioned inputs and requirements generated the prototypes shown in Figure 44 and Figure 45. The use of dark tones facilitated readability, especially the identification of yellow clusters. Also, filtering eliminated clusters that were considered irrelevant and collaborated to intense visual clutter.

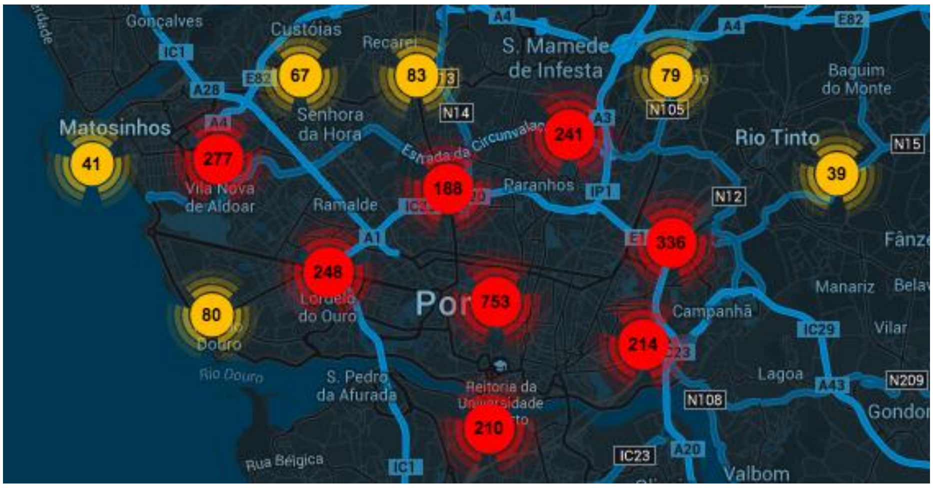


Figure 44 - Overview of clustered markers (monochromatic blue scale and filtered clusters)

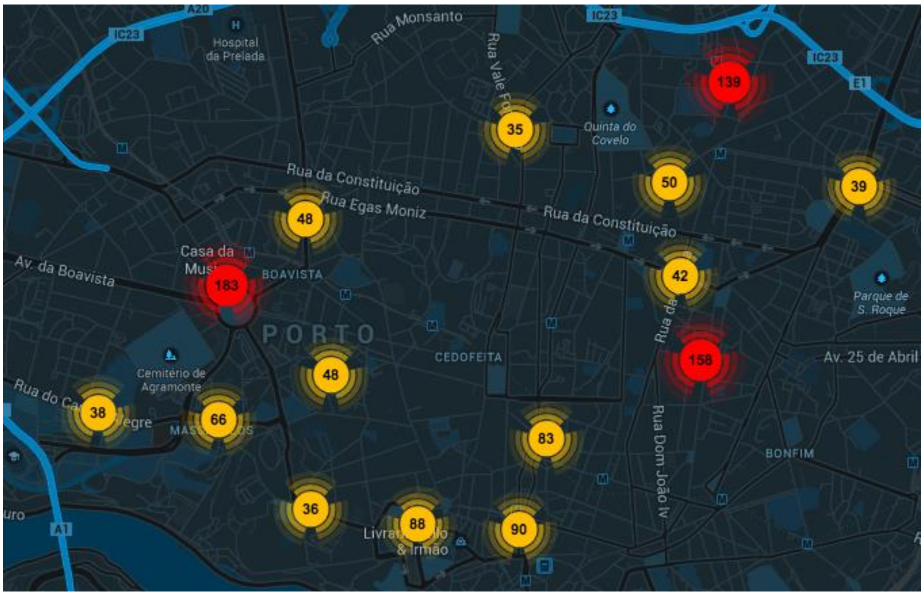


Figure 45 - Clustered markers focusing on downtown city zone

Visualisation 3 – Heat map

Readability: Minimal effort for understanding the visualisation concept. Experts expressed positive impressions about this visualisation, stating that it overcame the deficiencies detected in Circle markers. Experts were able to identify facts represented on the proposed examples.

Colour coding: The heat colour coding was considered adequate and intuitive. The standard map colour palette was considered distractive. The example that used a monochromatic palette was preferred.

Visual clutter: There were no mentions to visual clutter issue in this visualisation. Experts stated that the visualisation concept naturally tackles visual clutter successfully.

Suitability: Heat maps were considered a powerful way of analysing density of weighted spatiotemporal data, both at the level of the entire city (lower zoom levels) and at the level of a specific zone (higher zoom levels). However, they stated that heat maps are more effective if data belongs to a continuous or non-sparse discrete spatial domain, as in the case of traffic counter sensors data (see Figure 30) and Move-me's Near Stops service. If data consists of a small discrete set of sparse locations (see Figure 29), heat maps might not be equally effective. Experts believe that this visualisation might be significant for strategic decisions.

Requirements: The map colour palette should be monochromatic and considerably distinct from the colours that compose the heat colour scale.

The aforementioned inputs and requirement generated the prototypes shown in Figure 46 and Figure 47. Experts mentioned that dark maps are preferable, as grayscale maps may hinder the visualisation of green tones in the heat colour scale.

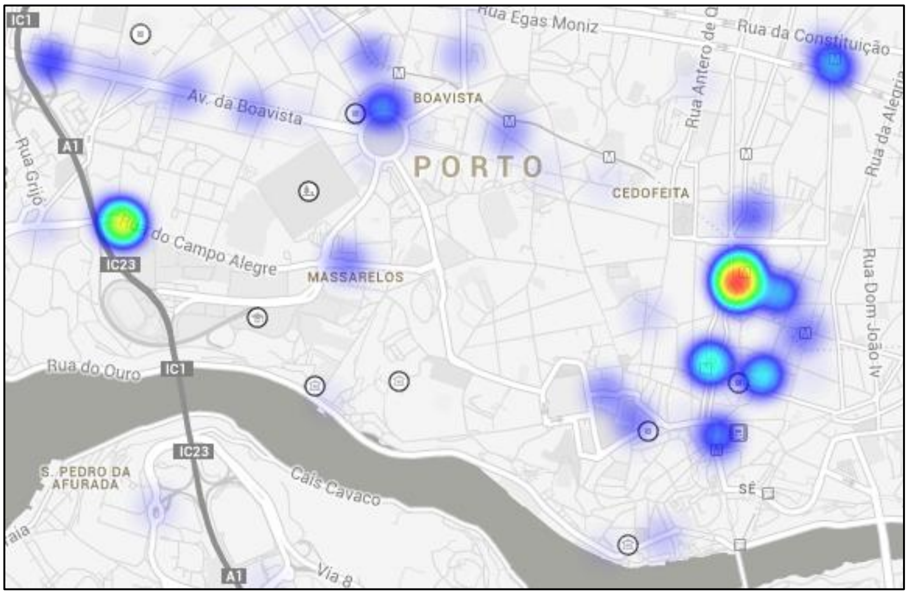


Figure 46 - Monochromatic map coupled with heat maps (grayscale)

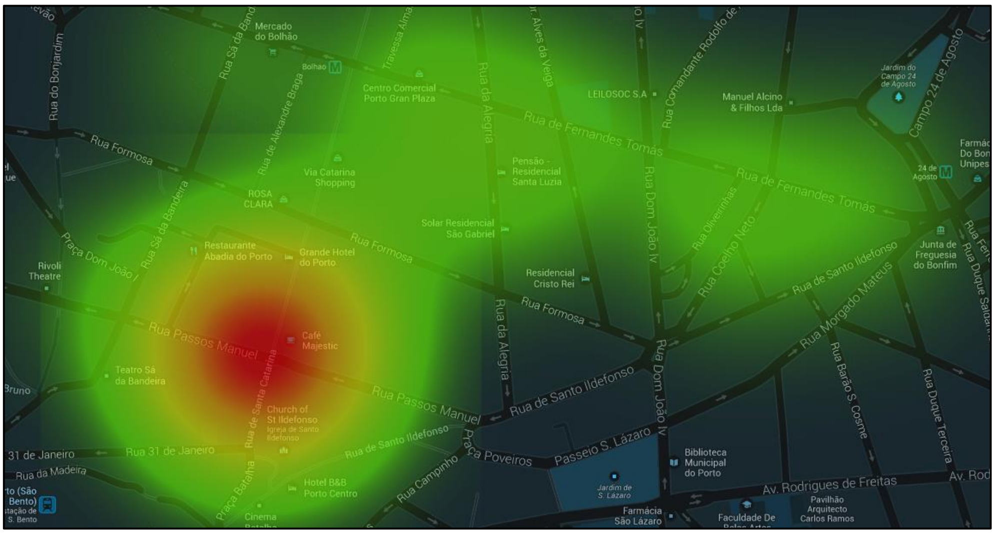


Figure 47 - Monochromatic map coupled with heat maps (blue scale)

Visualisation 4 – Sized circles

Readability: Small effort for understanding the visualisation concept, which was considered intuitive. Experts were able to identify facts and stated that circles clearly indicate extreme events.

Colour coding: Colour coding was considered consistent and mandatory. Experts rapidly identified each of the transport providers depicted in the visualisation.

Visual clutter: Clutter was considered a moderate issue in the case of having overlapped circles due to higher radius sizes. When they overlap, the user is unable to interact with the circles at are overlapped and retrieve specific information about it. This situation led experts to require workarounds such as filters in order to partially overcome this issue.

Suitability: Sized circles were considered suitable for analysing weighted data within a discrete spatial domain regardless of sparseness. The depiction of magnitude as size was considered important and adequate for identifying extreme and occasional events.

Requirements: The visualisation should allow filtering by transport provider according to an adjustable magnitude parameter.

Visualisation 5 – Calendar heat chart

Readability: Experts initially considered the visualisation concept difficult, and required additional instructions. There were difficulties to locate a specific day of the week and to extract knowledge due to the adopted heat scale. They stated that the colour scale should have no more than four steps. Extreme facts such as the national strike were promptly identified, as well as seasonal fluctuations for some stops.

Colour coding: Experts considered the colour coding inadequate and difficult to interpret. The increased number of heat levels combined with the monochromatic colour scales produced similar colours that could not be distinguished easily, especially when external factors were considered such as lighting and screen brightness.

Visual clutter: Colour coding and heat levels were the main causes of visual clutter. Also, day labels were considered distracting and compromised reading.

Suitability: Experts considered the visualisation powerful for analysing trends, fluctuations and anomalies on a yearly basis. They highlighted that it might not be useful only for visualising data for specific stops, but also aggregate data for multiple stops or city zones. Furthermore, they pointed that calendar heat charts would be significantly improved if other data sources (e.g. weather data) were integrated into it, as weather plays a significant role on mobility patterns according to their own experience. This visualisation would allow them to understand its impacts on mobility accurately. However, visual clutter issues were considered priority.

Requirements:

The month orientation should be horizontal; each row should correspond to a day of the week.

Day labels should be removed and should only appear when placing the cursor over it.

The aforementioned inputs and requirement generated the prototypes shown in Figure 48. Rather than organising days of the week in columns as in Figure 34, days are now organised in rows. The number of colour steps was reduced; colour coding was reworked and day labels only appear when the user places the cursor over a day.

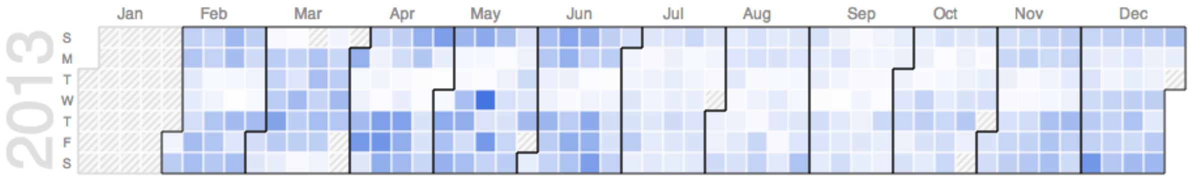


Figure 48 - Calendar heat chart prototype after experts' requirements

<u>Visualisation 6 – Line plots</u>

Readability: minimal effort for understanding the visualisation concept. Experts highlighted the importance of taking advantage of simple techniques that they are already familiar with. They pointed aspects to be considered depending on the data type:

For travel intentions data, lines should be smooth (Bézier curves). Experts justified that it is more important to analyse fluctuation than quantitative results when performing strategic analyses.

For ticketing data, when manipulating the visualisation of average validations for a specific line (Figure 36), experts considered that the intense oscillating pattern severely compromised readability and it did not bring useful information. For that visualisation, they suggested bar plots as a straightforward way to represent that data. That would be suitable for both strategic and operational analyses.

Colour coding: The contrast between foreground and background colours was considered adequate. Experts highlighted that simplicity should always be a priority in this type of visualisation, thus any shading or texture elements are distractive and irrelevant.

Visual clutter: Clutter was considered an issue either if multiple lines (more than 5) lines were plotted onto the same map, or if the density of ticks in the horizontal axis started stretching the line (e.g. in case of bus lines with a considerable amount of stops).

Suitability: Line plots were considered meaningful for visualising data from travel intentions, as they allow the prompt identification of extreme events and seasonality, and for being a powerful way to compare trends in different locations (e.g. stops) within the same visualisation. However, experts found line plots unsuitable for analysing ticketing data, as in our proposed visualisation, which they suggested to be changed to bar plots.

The issue with ticketing data motivated some experts to suggest another type of visualisation where line plots would then be suitable: *average accumulated validations for a specific line*, which would adequately represent the variation of validations along a bus line.

Requirements: Lines should be smooth when representing data from travel intentions or other trends.

Lines should be straight when representing data from travel events, and markers should highlight data points with their respective numerical values.

The aforementioned inputs and requirements generated the reworked prototype shown in Figure 49, which compares the use of lines and bars as proposed by experts.

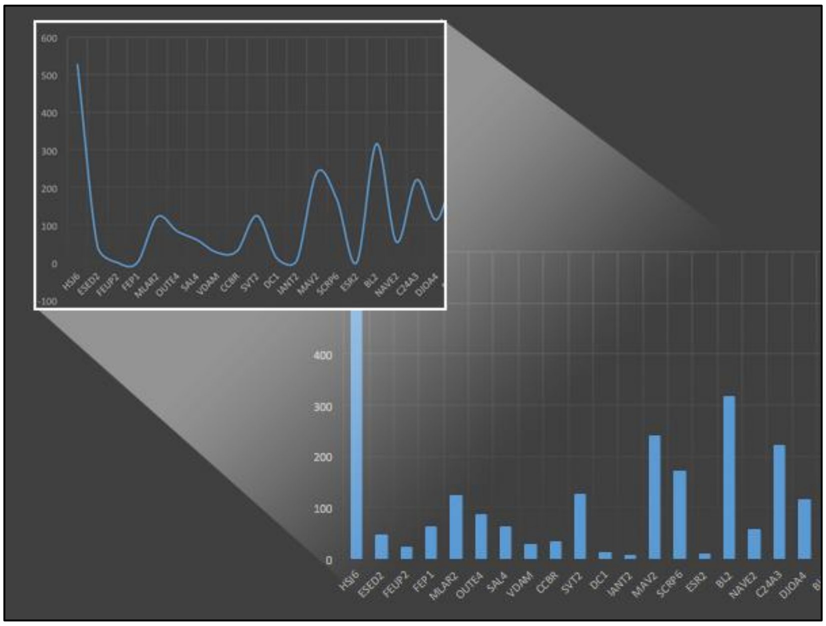


Figure 49 - Comparison between line and bar plots for the same data

Expert's suggestion for visualising *average accumulated validations for a specific line* was then represented on the following visualisation with line plots (see Figure 50). In that case, they considered that straight lines are more adequate, as they wanted to view the evolution of actual data without the distortions caused by Bézier curves due to the interpolation of data points. Significant variations in ridership levels are highlighted with yellow circles.

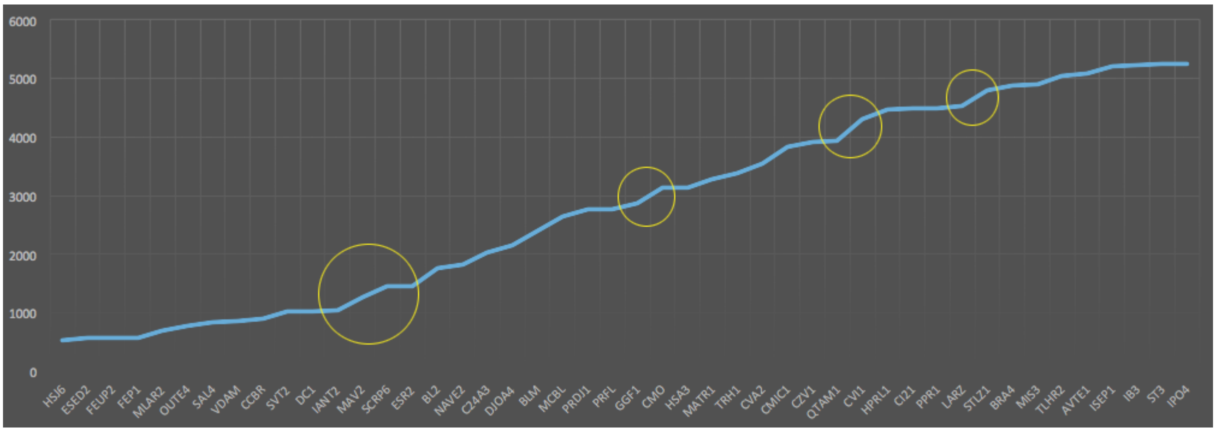


Figure 50 - Average accumulated validations for a specific line

Visualisation 7 – Radial heat chart

Readability: The visualisation concept was considered meaningful and a powerful way to represent large volumes of data into a compact graphic. Some experts highlighted that the choice for adopting radial axes was crucial for good readability.

Colour coding: Colour coding was considered consistent, although the increased number of colour steps led to some confusion, especially in light background colours. Experts pointed that 3-to-4 colour steps would be sufficient. In addition, experts asked for an additional example using the heat colour scale used in Heat Maps visualisation.

Visual clutter: Again, the amount of colour steps was considered prone to misinterpreting data.

Suitability: Experts considered radial heat chart a powerful visualisation for ticketing data that will help analysing seasonality in stops, and it might play a crucial role when supporting decisions involving changes in bus lines. However, this visualisation might not be equally meaningful when considering travel intentions, as experts believe this type of data might not be sufficiently significant for analyses from a weekly perspective.

Experts pointed that filters are very important, given that a stop might be served by more than one line, and the demand for each of them might imply separate decisions and reveal distinct patterns.

Requirements: The visualisation should allow filtering by specific lines.

Visualisation 8 – 3D Lines

Readability: The visualisation concept was considered straightforward, although experts had difficulties to identify convergence zones in Porto district due to the amount of lines.

Colour coding: Colour coding was considered adequate, having in mind the adopted map palette. Experts highlighted that darker maps prevented distractions.

Visual clutter: Visual clutter was considered an issue, although experts reckoned that when dealing with directed spatial data, it is inevitable to not have it somehow. They stated that the *strategy* of assigning low opacity to lines was valid for identifying convergence zones. However, they stated that some filtering is still required for this visualisation; otherwise it could lead to wrong identification of convergence zones that actually do not exist.

Again, experts emphasised that exact locations are irrelevant when analysing trends. Hence, lines could be clustered according to the relative position of their vertices, and magnitude

S: Despite the aforementioned issues, the visualisation was considered a complement for identifying extreme events and discovering new trends within the city. However, experts considered it experimental, given that visual clutter issues could lead to wrong interpretations. Clustering techniques would significantly improve this visualisation.

Requirements: The visualisation cluster lines in order to avoid visual clutter.

Visualisation 9 - Radial convergence layout

Readability: Initially, experts demonstrated some confusion about the visualisation concept. After demonstrating the interactive features of the visualisation, they were able to understand its concept successfully and were able to extract knowledge from it. They stressed that interaction is essential for making chord diagrams useful.

Colour coding: Experts stated that colour coding is not intuitive, although understandable.

Visual clutter: Visual clutter was promptly identified when the visualisation was displayed for the first time. However, they acknowledged that this visualisation is supposed to be interactive, and the interactive filtering successfully tackles the issue with visual clutter.

Suitability: Experts considered this visualisation a powerful tool for representing the flows of potential and actual passengers inside a transportation network. Experts considered the visualisation complex, and suggested that it could of better use for operational levels rather than strategic levels of transport related services and government authorities. They highlighted that it was a creative and functional way of representing data that is strongly connected to geographic visualisations.

Visualisation 10 – Geographic graph

Readability: Experts had no difficulties to understand the visualisation concept; although they spent some time trying to identify the direction of traffic flows between stops. After noticing that the visualisation consolidates flows from both directions, experts suggested that it would also be desirable to visualise flows for each direction. They pointed that the sizes of nodes should vary according to their respective inbound/outbound flows.

Colour coding: Dark background colours were considered adequate. The use of colours for distinguishing clusters of passengers was considered valuable for further in-depth studies.

Visual clutter: Visual clutter was considered an issue if many stops were plotted onto the visualisation or if there were a high amount of overlapped lines. They stressed the need of having filters in order to restrict the visibility of stops and curves according to adjustable parameters such as *flow magnitude* or *city zones*.

Suitability: Geographic graph was considered a powerful way for representing trends and actual passenger flows within a transportation network. Differently from the Radial convergence layout, experts stated that this visualisation should be aimed to strategic levels due to its intuitive concept.

Requirements: The visualisation should include filters for restricting the visibility of stops and curves according to *flow magnitude* or *city zones*.

The visualisation should allow choosing between showing consolidated flows and separate flows for each direction between two stops.

<u>Visualisation 11 – Coupled visualisations</u>

Readability: Experts were satisfied with the proposed coupling, and were able to extract knowledge and draw conclusions about the evolution of AADT throughout the considered period. Moreover, they stated that the visualisation allowed the immediate recognition of an extreme event at "Magalhães (4)" sensor.

Colour coding: The map colour palette was considered adequate and offered no distractions. Colour coding for AADT variation was also considered adequate, although they pointed that red/blue might be interpreted as negative and positive values respectively.

Visual clutter: There were no potential occurrences of visual clutter, although they pointed that some graphic elements could be hidden initially. For example, it was not necessary to show name labels for all traffic counter sensors, as experts already know their location precisely. Instead, a label should appear whenever the user places the cursor over it.

Suitability: This combination of visualisations was considered powerful for analysing variations or other pre-defined variables based on historical data. Experts mentioned that they could suit analyses at both strategic and operational levels, and could be integrated with ticketing and travel intentions data. In addition, they pointed that interaction could be improved if line plots appeared immediately when placing the mouse over a circle, instead of only showing them after clicking.

Requirements:

The label name for a circle should only appear when the cursor is placed over it.

Line plots should appear when the cursor is placed over a circle.

Visual query control

Experts stated that the query control should feature advanced options for querying data that allow one to manipulate days and hours/minutes separately, instead of date/time instances as in our current query control. Also, they wanted to query data according to specific days of the week and facts (e.g. holidays, city events).

Experts' reactions to visualisations

Besides the opinions and requests made by experts, we observed subjective and other reactions during usability tests. During the presentation of prototypes, experts usually engaged in discussions between them, trying to understand the leading causes for the facts expressed on visualisations. Subjective reactions consisted on some level of excitement externalised by experts themselves when presented to some visualisations, mainly in heat maps, radial convergence layout, calendar heat charts and radial heat charts.

Some experts placed assertive conclusions from some of the facts expressed in visualisations. For example:

> *"This is one of the proofs that Aliados station was unnecessary. Trindade and São Bento are able to handle the existing demand."*
>
> An expert's comment on the information derived from the visualisation in Figure 33

Experts stated that visualisations even offered some level of excitement, which some of them expressed as "some sort of happiness and satisfaction", given that it was possible to visualise data from different perspectives that they still do not make use of. When asked about the reason of not using visualisation techniques, they emphasised the inherent technological difficulties and time constraints for manipulating raw data and "translating" them into visual representations.

Despite the availability of visualisation tools and Geographic Information Systems that do not require in-depth programming knowledge from users, they stated that it is still difficult to take full advantage of those technologies in their context, and most of the decision-support activities only consisted in manipulating Excel worksheets. Finally, they pointed that it is also difficult to find professionals with substantial proficiency in those tools and experience with transport and urban mobility.

Chapter 6 Discussion

The goal of this chapter is to interpret our findings about the adopted User-Centred Design process and the resulting visualisation prototypes, aligning them with the topics regarded in the literature review.

6.1 The User-Centred Design process

The exploratory usability tests with experts occurred as expected. Considering the natural limited availability of experts, the timeframe adopted for the test sessions was adequate and sufficient for carrying out all the intended activities.

We were concerned about well-acknowledged usability principles described in Chapter 2. The *early focus on users and tasks* allowed us to understand what are the essential characteristics to be considered when developing visualisations for urban mobility. Although we considered a preliminary phase for analysing the context of use, we found that every session with experts revealed additional information about how could visualisations be applied on their daily routines. This is explained by the participation of experts with significant knowledge of urban mobility, but with different backgrounds and activities within this domain. We reckon that different but complementary perspectives are valuable for exploratory studies in Information Visualisation. This implied that the Context of Use analysis phase should be constantly revisited during the UCD cycle, rather than being considered as a preliminary phase as we initially suggested, which leads to the following change in relation to our UCD process:

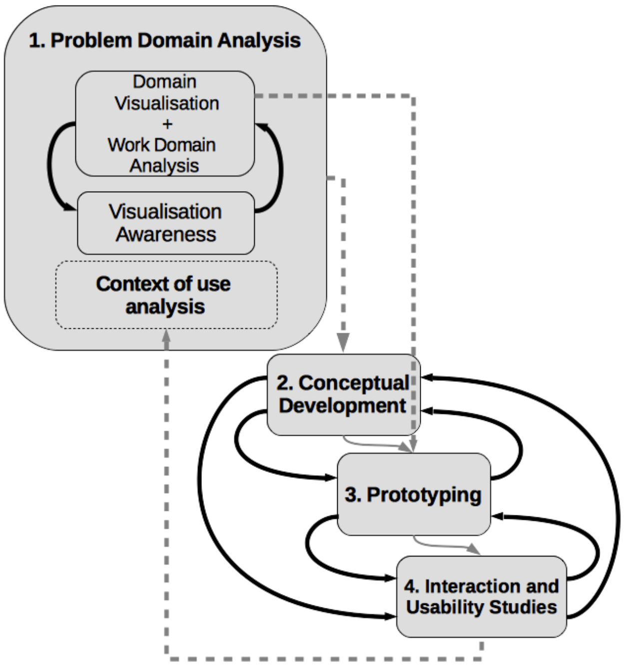


Figure 51 - The UCD process adopted in this research, with changes regarding the Context of use analysis phase

The *constant evaluation with experts* and further *design iterations* allowed us to align visualisations with their needs and preferences. Regarding other UCD approaches mentioned in Chapter 2, we believe that they would not provide the same richness of results that we were able to come up with. In fact, linear approaches as the waterfall model would not be adequate to our research, considering its inherent exploratory nature. Moreover, the new findings from each meeting were only made possible due to the constant iteration of all phases of the UCD process, which significantly differs in nature from the waterfall model. We stress the need of considering evolutional approaches when developing visualisations.

Furthermore, we acknowledged the issues stated by Ellis and Dix (2006) regarding the difficulties in evaluating visualisations:

Regarding the *variety of datasets*, we wanted to effectively overcome the difficulties in dealing with heterogeneous data. The analysis of datasets followed by a formal data characterisation and time modelling were fundamental for allowing us to systematically understand the data we had available and avoid revisiting data handling phases.

The *indeterminacy of tasks* was not a significant difficulty in this research, considering our focus on exploring the potential benefits and applications of our visualisations rather than evaluating them with structured tasks, which would better suit later advances in this research.

The *participants in context* were fundamental for ensuring the adequacy of our results. We believe that exploratory visualisation studies should prioritise approaching work domain experts whenever possible. This is justified for the constant iterations of Work Domain phases. Moreover, the discussions that we observed between experts during the test sessions were a positive confirmation that visualisations could successfully promote moments of reflection about facts.

We conclude that, if possible, experts with different perspectives of urban mobility should be involved. We experienced a few difficulties during the first meetings when asking experts to also evaluate visualisations at a conceptual level, i.e. abstracting from the concrete visualisation examples we presented. In order to facilitate this task, we integrated prototypes with other types of data in order to help experts to abstract. We initially expected that the free manipulation of prototypes would collaborate with that purpose. However, our experiences lead us to infer that it actually produces more concrete implications for interaction and usability aspects rather than the conceptual interpretation of visualisations.

Finally, the adopted UCD process was aligned with our research purposes and its outcomes are valid within its scope. Considering the qualitative impressions collected from experts, we conclude that the accepted visualisations positively contribute to the achievement of experts' goals within the context of use of visualisations. Although this research does not cover implementation phases, we consider that the UCD methodology proposed by Lian Chee *et al.* (2011) provide satisfactory outcomes in the urban mobility domain.

6.2 Visualisation prototypes

6.2.1 Geographic visualisations

Geographic visualisations typically received positive feedbacks. Experts' impressions and requirements concentrate on filtering and colour coding issues; concretely the preference for dark and monochromatic colour palettes on maps. Considering that our suggested colour coding for data typically featured different colours, it is sensible that polychromatic map colour palettes contribute to visual clutter.

Before evaluation, we believed that polychromatic maps would still be desirable, as they offered visual cues about different areas and elements of a city. However, experts pointed that they are already familiarised with the typical dynamics of each city zone, which eliminates the need of adopting detailed maps beforehand.

Heat maps are adequate for representing data from travel intentions, travel events and traffic flows. Heat maps suit strategic-level analyses at both macro and micro geographic areas, and are more effective when representing weighted data within a continuous spatial domain. For discrete spatial domains, spatial sparseness can compromise the utility of heat maps, thus it is preferable to make use of discrete spatial domains with low sparseness such as data from traffic flows.

Clustered markers are adequate for representing data from travel intentions within a continuous spatial domain, and are suitable for operational and strategic-level analyses, given the emphasis on providing numerical indications based on the density of data primitives over an area. However, it is important to introduce filters for adjusting clusters and hiding unnecessary small clusters that cause visual clutter.

Sized circles are adequate for representing data from travel intentions, travel events and traffic flows within a discrete spatial domain, regardless of sparseness. This visualisation was considered to overcome the deficiencies of heat maps when representing data within discrete and sparse spatial domains.

Geographic Graph is adequate for representing data from travel intentions and ticketing data within a directed spatial domain, and promotes rapid recognition of trends along the transport network. Compared with the Radial convergence layout mentioned in Section 6.2.2, Geographic Graph is more suitable for strategic analyses. Filtering is essential for avoiding visual clutter and for visualising flows in different directions between two stops.

We conclude that the aforementioned visualisations satisfy the usability measures described in Section 2.1.1. Effectiveness and satisfaction benefitted from the rapid learnability of our visualisation concepts. Requirements and suggestions about issues found in visualisations should be addressed so as to increase efficiency and reduce the risk of misinterpreting data.

Circles visualisation did not satisfy usability measures, as it failed to represent the notion of density within a continuous spatial domain, and it was considered irrelevant given the availability of other visualisations that were approved. Although 3D Lines had its potential recognised by experts, it should still be improved in order to properly satisfy usability measures and effectively support knowledge extraction. Experts considered that lines are more suitable than arcs, and the latter are more suitable when representing data within a spatial domain that covers a broad geographic region such as countries or continents. However, the techniques used for representing data still cause intense visual clutter that may cause data misinterpretation, which should be addressed by implementing a type of clustering algorithm.

6.2.2 Abstract visualisations

Calendar heat charts are adequate for representing data from travel intentions and travel events for a specific entity (e.g. stop) that can be assigned to a weight according to a specific parameter (e.g. amount of schedule requests). It is particularly suitable for strategic analyses, focusing on detecting seasonality and extreme events. Considering that each data point represents an indication for a day of the year, the colour coding scale should have few steps (not more than 4) in order to facilitate reading. We reckon that the experts' requirements for removing day labels and changing the calendar orientation improved readability and eliminated visual clutter.

Line plots are adequate for representing data from travel intentions and travel events. For each of those types of data, different line styles were preferred. For travel intentions, which are usually considered for strategic analysis, it is desirable to represent data with Bézier curves instead of straight lines. Regarding travel events, more specifically our visualisation proposal (see Figure 36), experts suggested the use of bar plots. We concluded that although our goal was to represent the amount of validations along each stop and still introduce the notion of variation along the line, experts in that case were more concerned with the individual values for each stop.

Radial heat charts are adequate for representing data from travel events, for both strategic and operational analyses. Experts demonstrated particular satisfaction with this visualisation, as they considered that it is able to minimize reading effort and still represent a considerable amount of data meaningfully. This re-confirms the earlier results in existing literature regarding radial layouts (see p. 40). As in Calendar heat charts, the number of colour steps should be limited.

Radial convergence layout diagrams are adequate for representing data from travel intentions and travel events between stops of a transportation network. This visualisation was considered adequate for operational-level analyses due to its complexity. In fact, professionals belonging to strategic levels typically want to extract knowledge as fast as possible, and in those situations the Geographic Graph shows to be more adequate.

6.2.3 Coupled visualisations

Regarding experts' positive comments about coupled visualisations, more concretely about the coupled prototype we proposed, we conclude that the combination of geographic with abstract visualisations should be explored, as it can offer new perspectives for users to interpret and extract knowledge from data. Acknowledging Shneiderman's recommendations (1996), we believe that geographic visualisations can support in tasks such as *overview* and *zooming*. Abstract ones can support users when requesting *details on demand*, after *filtering* data according to their preferences and needs. This poses new interaction and usability challenges on how to make simultaneous use of different visualisation types in order to improve cognition without compromising readability and introducing graphic artefacts that can lead to distraction.

Chapter 7 Conclusions and evolution perspectives

This research proposed a thorough user-centred design methodology for developing visualisations for urban mobility data. It intended to emphasise the importance of involving users throughout the design process and the benefits of rapid prototyping. It was crucial to elaborate functional prototypes that made use of real urban mobility data, so as to ensure that urban mobility experts were able to extract knowledge from our visualisation proposals.

Prior to the development of prototypes, we performed a systematic analysis of the datasets in order to understand the nature of data contained into them, as well as to identify preliminary data adaptations that would be needed before proceeding. This reaffirms the underlying challenges of Information Visualisation, which deals with heterogeneous datasets that come from different sources. The analysis also considered a formal characterisation of data. After developing an initial set of visualisation prototypes, we held exploratory usability tests with experts that comprised free manipulation of prototypes and semi-structured interviews. Experts' statements and impressions ignited the iterative development of prototypes. Finally, the results allowed identifying which visualisations were more adequate for extracting knowledge, according to the data type that is concerned.

7.1 Research questions revisited

RQ1. *Can the proposed user-centred design methodology produce meaningful visualisations for the urban mobility domain?*

The UCD methodology adopted in this research was able to produce meaningful visualisations that are in accordance with the needs of urban mobility experts. This alignment results from the active involvement of work domain experts during all phases. The engagement of experts with different but complementary perspectives on urban mobility allowed us to gain in-depth understanding of their work domain, which reflected on the adequacy of visualisation prototypes according to usability measures.

We highlight the importance of revisiting the Context of Use analysis phase during the UCD process instead of keeping it as a preliminary phase, as our own experience showed that new potential applications of visualisations in the urban mobility domain appeared at every test session with experts.

Furthermore, we acknowledged that it might be difficult for experts to abstract from the visualisation prototypes that they are being presented to. In order to overcome this difficulty, we offered additional examples by integrating different types of data to our prototypes, which offered some facilitation.

RQ2. *Depending on the data that is concerned, which visualisation techniques support knowledge extraction and collaborate on decision-making routines of urban mobility experts?*

Based on the results of the UCD process, we were able to classify the visualisation techniques according to the *type of urban mobility data* that is concerned, and other characteristics: *frame of reference*, *spatial domain*, *weight* and *scope* (see Table 7).

Table 7 - Classification of visualisation techniques

Visualisation	Frame of reference		Type of urban mobility data			Spatial domain			Weight		Scope	
	Geographic	Abstract	Travel intentions	Travel events	Traffic flows	Discrete	Continuous	Directed graph	Weighted	Non-weighted	Strategic	Operational
Clustered markers	●		●				●			●		●
Heat map	●		●	●	●	●	●		●	●	●	
Sized circles	●		●	●	●	●			●		●	●
Calendar heat chart		●	●	●		●			●		●	
Line plots		●	●	●		●	●		●	●	●	●
Radial heat chart		●		●		●			●		●	●
3D Lines (*experimental*)	●			●				●	●	●	●	
Radial convergence layout		●	●	●				●	●			●
Geographic Graph	●		●	●				●	●	●	●	
Coupled Visualisation	●	●	●	●	●	●			●		●	●

7.2 Research limitations and evolution perspectives

Considering that the proposed visualisations are to be implemented on a future stage, we strived to approach a significant group of urban mobility experts from a variety of segments such as transport companies and government authorities. Also, this was done so in order to compensate the fact that involving experts on user-centred studies is difficult due to their limited availability. Although this work is focused on exploring experts' impressions about certain visualisation methods, an extended availability would have allowed the application of more in-depth usability evaluation methods such as task scenarios.

From the data perspective, most of the proposed visualisation methods focus on the "stop" level of the transportation network, which means that most visualisations have stops as the core element for analysis. Other perspectives such as "line" and "vehicle" level are also fundamental and should be regarded in the future, as they pose additional complexity and new challenges for finding adequate ways for visually representing data at those levels.

Our approach for visualising the time dimension consists of a static facet, i.e. a “snapshot” of a given time interval. Our visualisations did not allow experts to browse through time continuously like a timeline, for example. Future studies should incorporate more sophisticated time controls within the visualisation prototypes and assess experts’ impressions about the impact and suitability of time controls within the proposed visualisations. Advanced time controls fundamentally imply the need of adopting robust temporal query languages in order to minimise the complexity of time-based queries and to allow more refined results.

Regarding evolutional perspectives, we consider performing a deeper characterisation of data by taking advantage of existing approaches on the development of ontologies for transport. In the long run, we consider developing a visualisation tool aligned with the needs of urban mobility experts that would feature the visualisation proposals developed during this research.

References

Adrienko, Natalia, Gennady Adrienko, and Peter Gatalsky. 2005. *Impact of Data and Task Characteristics on Design of Spatio-Temporal Data Visualization Tools*. Edited by J. Dykes, A.M. MacEachren and M.-J. Kraak, *Exploring Geovisualization (International Cartographic Association).*

Aigner, Wolfgang, Silvia Miksch, Heidrun Schumann, and Christian Tominski. 2011. *Visualization of Time-Oriented Data*. 2011 ed, *Human-Computer Interaction Series*: Springer.

Benyon, David. 2010. *Designing Interactive Systems A Comprehensive Guide to HCI and Interaction Design (2nd Edition)*: Pearson Education Canada.

Bevan, Nigel. 2009. "International Standards for Usability Should Be More Widely Used." *Journal of Usability Studies* 4 (3):106-113.

Boehm, B. W. 1988. "A spiral model of software development and enhancement." *Computer* 21 (5):61-72. doi: 10.1109/2.59.

Christopher, Robert Kosara, Christopher G. Healey, Victoria Interrante, David H. Laidlaw, and Colin Ware. 2003. "Thoughts on User Studies: Why, How, and When." *IEEE Computer Graphics and Applications* 23:2003.

Du, Fan, Joshua Brulé, Peter Enns, Varun Manjunatha, and Yoav Segev. 2013. MetroViz: Visual Analysis of Public Transportation Data.

Earthy, J. 1998. Usability Maturity Model: Human Centredness Scale. Information Engineering Usability Support Centres.

Ellis, Geoffrey, and Alan Dix. 2006. "An explorative analysis of user evaluation studies in information visualisation." Proceedings of the 2006 AVI workshop on BEyond time and errors: novel evaluation methods for information visualization, Venice, Italy.

Gould, John D., and Clayton Lewis. 1985. "Designing for usability: key principles and what designers think." *Commun. ACM* 28 (3):300-311. doi: 10.1145/3166.3170.

Gulliksen, Jan, Bengt Göransson, Inger Boivie, Jenny Persson, Stefan Blomkvist, and Åsa Cajander. 2005. "Key Principles for User-Centred Systems Design." In *Human-Centered Software Engineering — Integrating Usability in the Software Development Lifecycle*, edited by Ahmed Seffah, Jan Gulliksen and MichelC Desmarais, 17-36. Springer Netherlands.

Hägerstraand, Torsten. 1970. "WHAT ABOUT PEOPLE IN REGIONAL SCIENCE?" *Papers in Regional Science* 24 (1):7-24. doi: 10.1111/j.1435-5597.1970.tb01464.x.

ISO 9241-11. 1998. Ergonomic requirements for office work with visual display terminals (VDTs) - Part 11: Guidance on usability.

ISO 9241-210. 2008. Human-centred design processes for interactive systems.

Jokela, Timo, Netta Iivari, Juha Matero, and Minna Karukka. 2003. "The standard of user-centered design and the standard definition of usability: analyzing ISO 13407 against ISO 9241-11." Proceedings of the Latin American conference on Human-computer interaction, Rio de Janeiro, Brazil.

Karat, John. 1996. "User centered design: quality or quackery?" *interactions* 3 (4):18-20. doi: 10.1145/234813.234814.

Karat, John. 1997. "Evolving the scope of user-centered design." *Commun. ACM* 40 (7):33-38. doi: 10.1145/256175.256181.

Keim, D. A., F. Mansmann, J. Schneidewind, and T. Schreck. 2006. "Monitoring Network Traffic with Radial Traffic Analyzer." Visual Analytics Science And Technology, 2006 IEEE Symposium On, Oct. 31 2006-Nov. 2 2006.

Kraak, Menno-Jan, and Alexandra Koussoulakou. 2005. "A Visualization Environment for the Space-Time-Cube." In *Developments in Spatial Data Handling*, 189-200. Springer Berlin Heidelberg.

Lian Chee, Koh, A. Slingsby, J. Dykes, and Kam Tin Seong. 2011. "Developing and Applying a User-Centered Model for the Design and Implementation of Information Visualization Tools." Information Visualisation (IV), 2011 15th International Conference on, 13-15 July 2011.

Lorenzo, Giusy di, Marco Luca Sbodio, Francesco Calabrese, Michele Berlingerio, Rahul Nair, and Fabio Pinelli. 2014. "AllAboard: visual exploration of cellphone mobility data to optimise public transport." Proceedings of the 19th international conference on Intelligent User Interfaces, Haifa, Israel.

Mack, Robert, and Jill Burdett Robinson. 1992. "When novices elicit knowledge: question asking in designing, evaluating, and learning to use software." In *The psychology of expertise*, 245-268. Springer-Verlag New York, Inc.

Mennis, Jeremy L., Donna J. Peuquet, and Liujian Qian. 2000. "A Conceptual Framework for Incorporating Cognitive Principles into Geographical Database Representation." *International Journal of Geographical Information Science* 14 (6):501-520.

Nielsen, Jakob. 1993. *Usability Engineering*: Morgan Kaufmann Publishers Inc.

Norman, Donald A., and Stephen W. Draper. 1986. *User Centered System Design; New Perspectives on Human-Computer Interaction*: L. Erlbaum Associates Inc.

Polisciuc, Evgheni, Ana Alves, Carlos Bento, and Penousal Machado. 2013. "Visualizing Urban Mobility." Special Interest Group on Computer Graphics and Interactive Techniques Conference, Anaheim, California.

Preece, Jenny, Yvonne Rogers, and Helen Sharp. 2002. *Interaction Design*: John Wiley & Sons, Inc.

Robinson, Anthony C., Jin Chen, Eugene J. Lengerich, Hans G. Meyer, and Alan M. Maceachren. 2005. "Combining Usability Techniques to Design Geovisualization Tools for Epidemiology." *Information Science* 32:243-255.

Roth, Robert E, Kevin S Ross, Benjamin G Finch, Wei Luo, and Alan M MacEachren. 2010. "A user-centered approach for designing and developing spatiotemporal crime analysis tools." *Zurich, Switzerland*.

Sagl, Günther, Martin Loidl, and Euro Beinat. 2012. "A Visual Analytics Approach for Extracting Spatio-Temporal Urban Mobility Information from Mobile Network Traffic." *ISPRS International Journal of Geo-Information* 1 (3):256-271.

Shneiderman, Ben. 1996. "The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations." Proceedings of the 1996 IEEE Symposium on Visual Languages.

Shneiderman, Ben. 1998. *Designing the user interface: strategies for effective human-computer interaction*. 3rd ed: Addison-Wesley Longman Publishing Co., Inc.

Shneiderman, Ben, Jock D. Mackinlay, and Stuart K. Card, eds. 1999. *Readings in information visualization: using vision to think*. Edited by Stuart K. Card, Jock D. Mackinlay and Ben Shneiderman: Morgan Kaufmann Publishers Inc.

Slocum, Terry A., Daniel C. Cliburn, Johannes J. Feddema, and James R. Miller. 2003. "Evaluating the Usability of a Tool for Visualizing the Uncertainty of the Future Global Water Balance." *Cartography and Geographic Information Science* 30 (4):299-317. doi: 10.1559/152304003322606210.

Spence, Robert. 2007. *Information Visualization: Design for Interaction (2nd Edition)*: Prentice Hall.

Tory, M., and T. Moller. 2005. "Evaluating visualizations: do expert reviews work?" *Computer Graphics and Applications, IEEE* 25 (5):8-11. doi: 10.1109/MCG.2005.102.

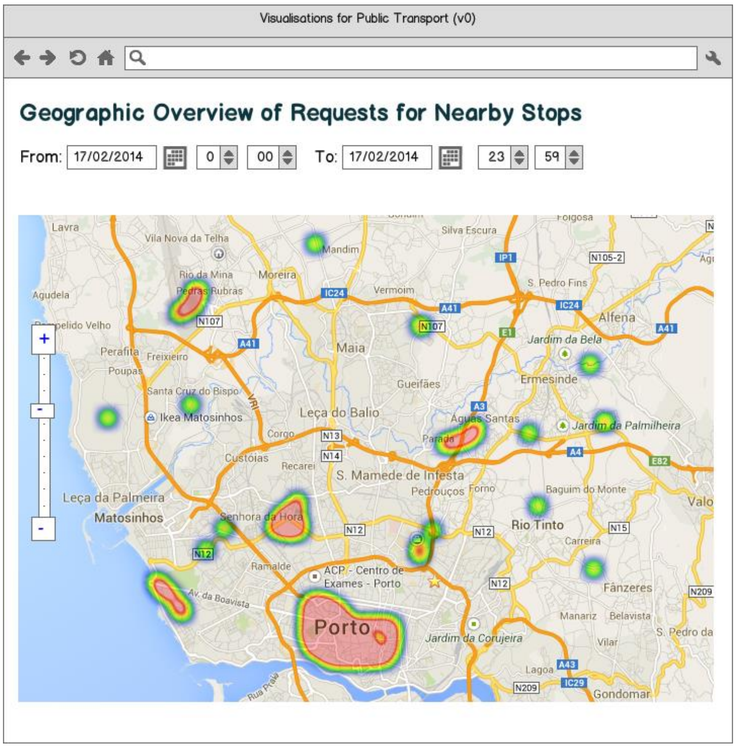
United Nations. 2012. World Urbanization Prospects - The 2011 Revision. New York: United Nations.

Ward, Matthew, Georges Grinstein, and Daniel Keim. 2010. *Interactive Data Visualization: Foundations, Techniques, and Applications*: A. K. Peters, Ltd.

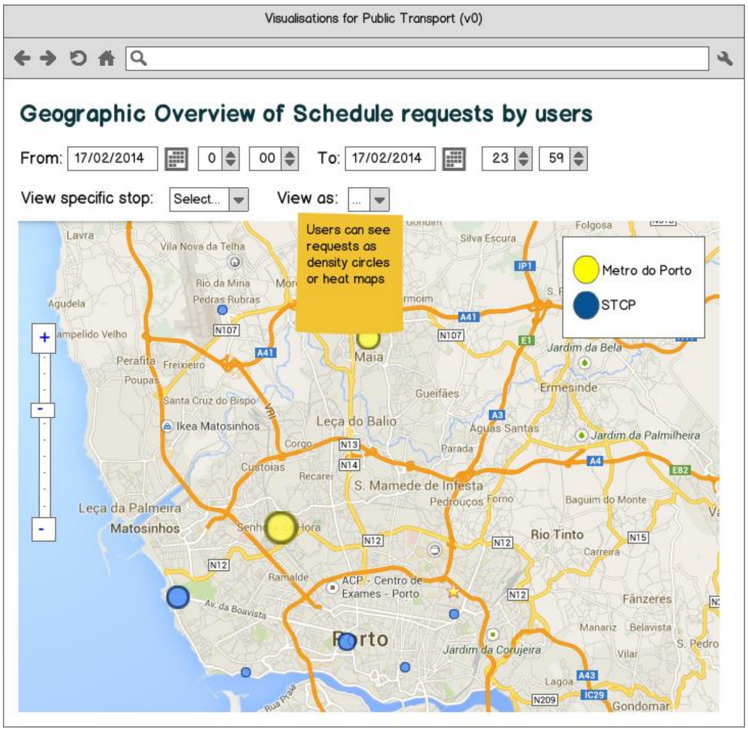
Ware, Colin. 2004. *Information Visualization: Perception for Design*: Morgan Kaufmann Publishers Inc.

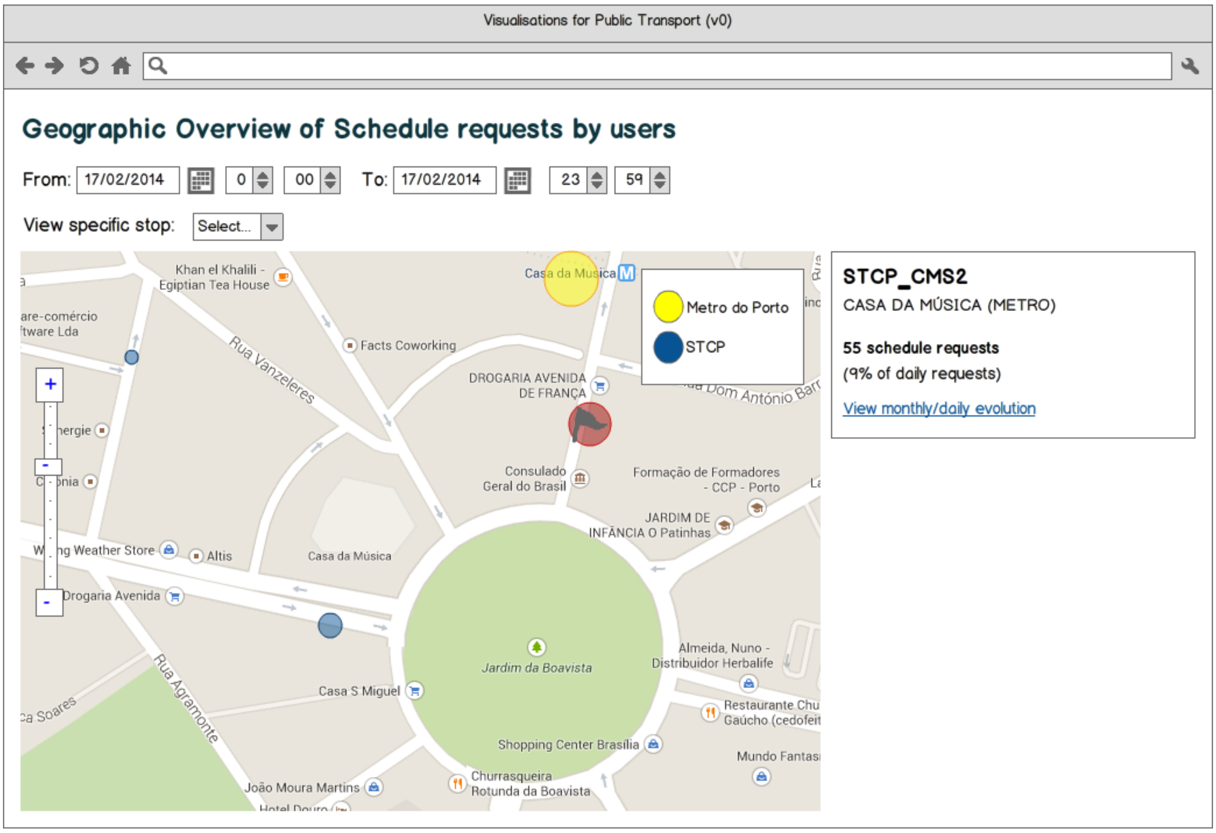
APPENDIX A: Examples of initial mock-ups from Domain Visualisation phase

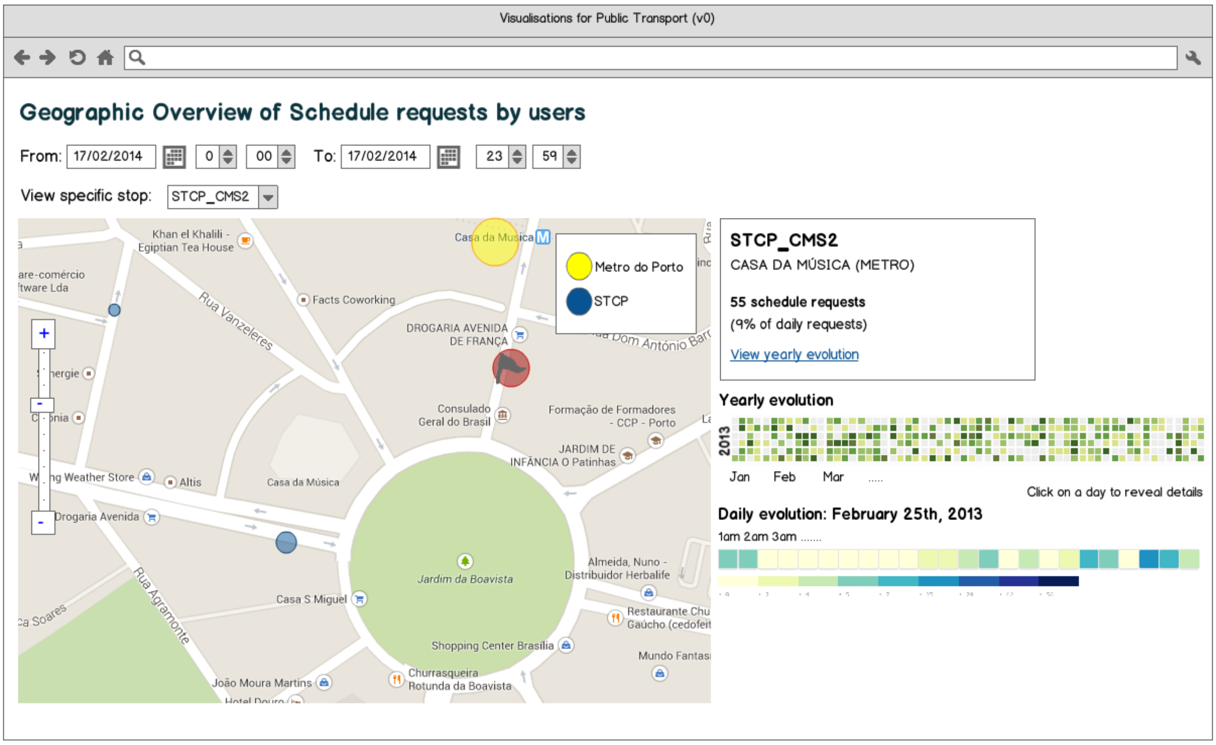
Near stops service – Heat map



Next schedules service – Density circles




Visualisations for Public Transport (v0)
Geographic Overview of Schedule requests by users
From: 17/02/2014 0 00 To: 17/02/2014 23 59
View specific stop: Select...
Metro do Porto
STCP
STCP_CMS2
CASA DA MÚSICA (METRO)
55 schedule requests
(9% of daily requests)
View monthly/daily evolution
Jardim da Boavista
Rua Vanzeleres
Rua Agramonte
Casa da Música
Shopping Center Brasilia


Visualisations for Public Transport (v0)
Geographic Overview of Schedule requests by users
From: 17/02/2014 0 00 To: 17/02/2014 23 59
View specific stop: STCP_CMS2
Metro do Porto
STCP
STCP_CMS2
CASA DA MÚSICA (METRO)
55 schedule requests
(9% of daily requests)
View yearly evolution
Yearly evolution
2013
Jan Feb Mar .....
Click on a day to reveal details
Daily evolution: February 25th, 2013
1am 2am 3am ....
Jardim da Boavista
Rua Vanzeleres
Rua Agramonte
Casa da Música
Shopping Center Brasilia

APPENDIX B: Guide for semi-structured interview (Domain Visualisation phase)

A) The visualisation method

1) Do you think this visualisation method is useful considering the data it represents? Why?

2) Is the information presented in a meaningful way? If not, why? Which improvements could be made?

3) What do you think about the adopted colour pallets, shapes and layout?

B) Knowledge extraction from visualisations

4) Which facts could be identified here? (*in this case, "here" refers to the visualisation that is being presented, which is showing data for a specific city zone in a given day/time).*

C) Visual query tool

5) Is the visual query tool adequate? How could it be improved?

D) Combining visualisations

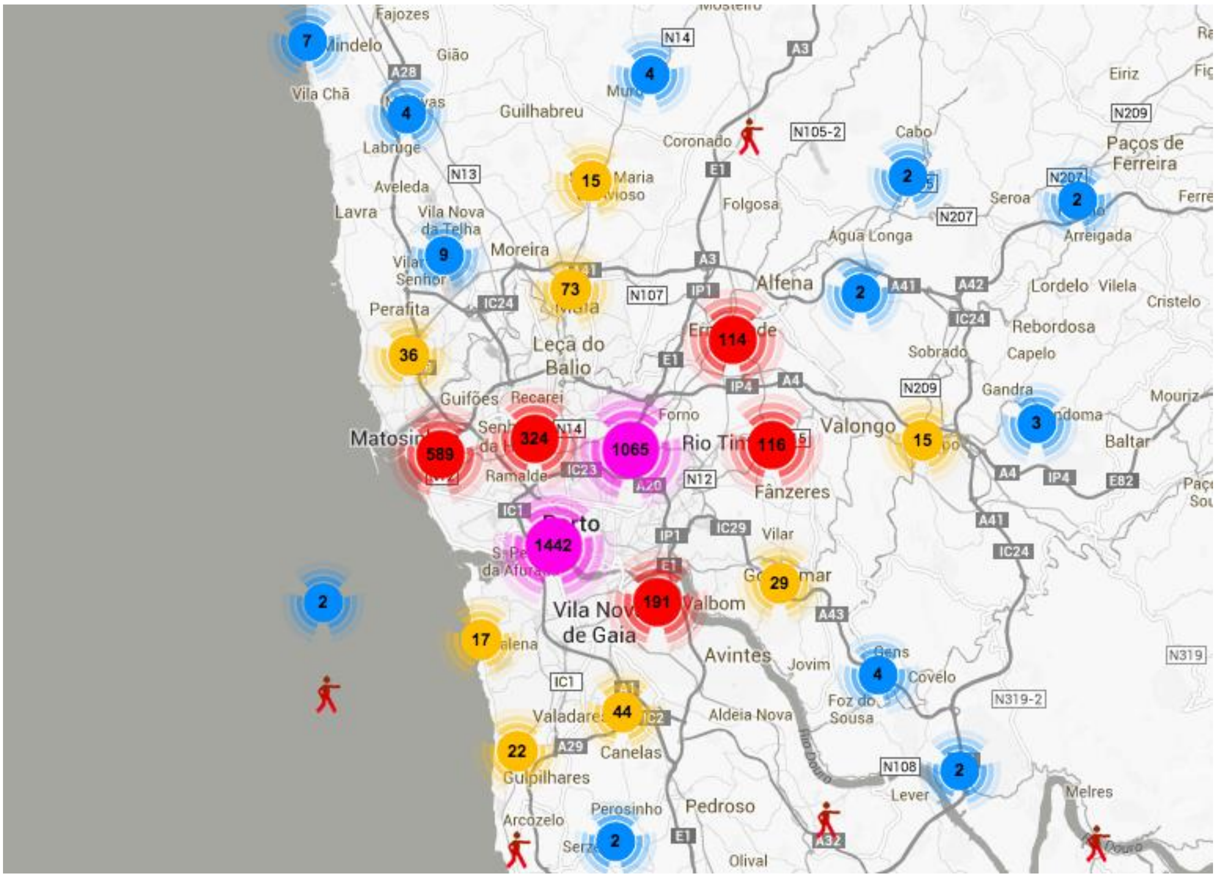
6) Do you think that some of the visualisations would benefit from combination? Which of them?

7) If you could rank the visualisations according to its characteristics and applicability to the transport domain, how would it be?

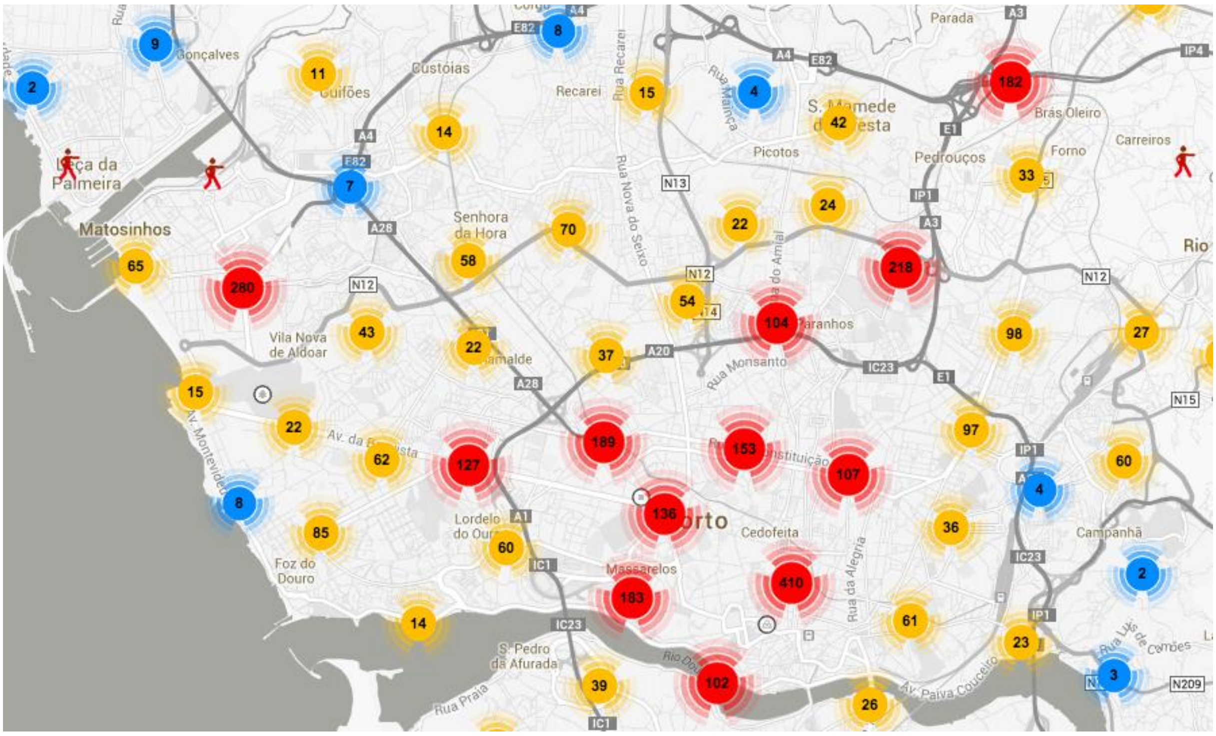
APPENDIX C: High-resolution captures of visualization prototypes

Clustered markers

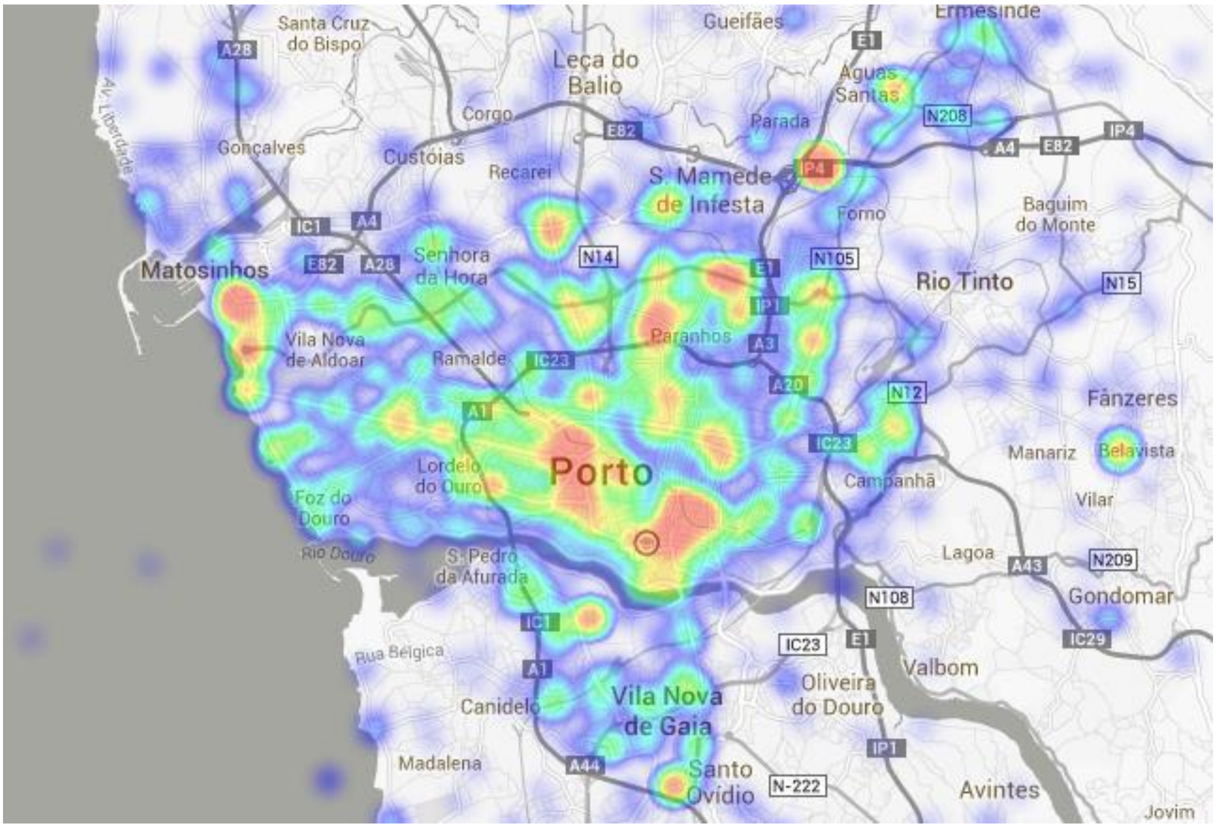
Initial zoom level

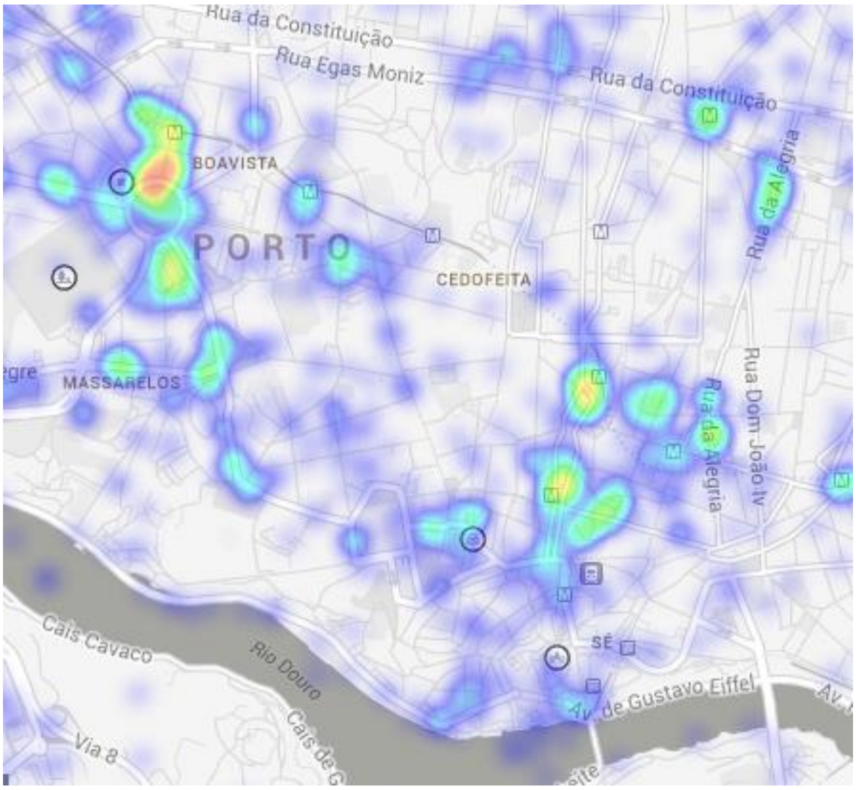


Zoom in (1 step)

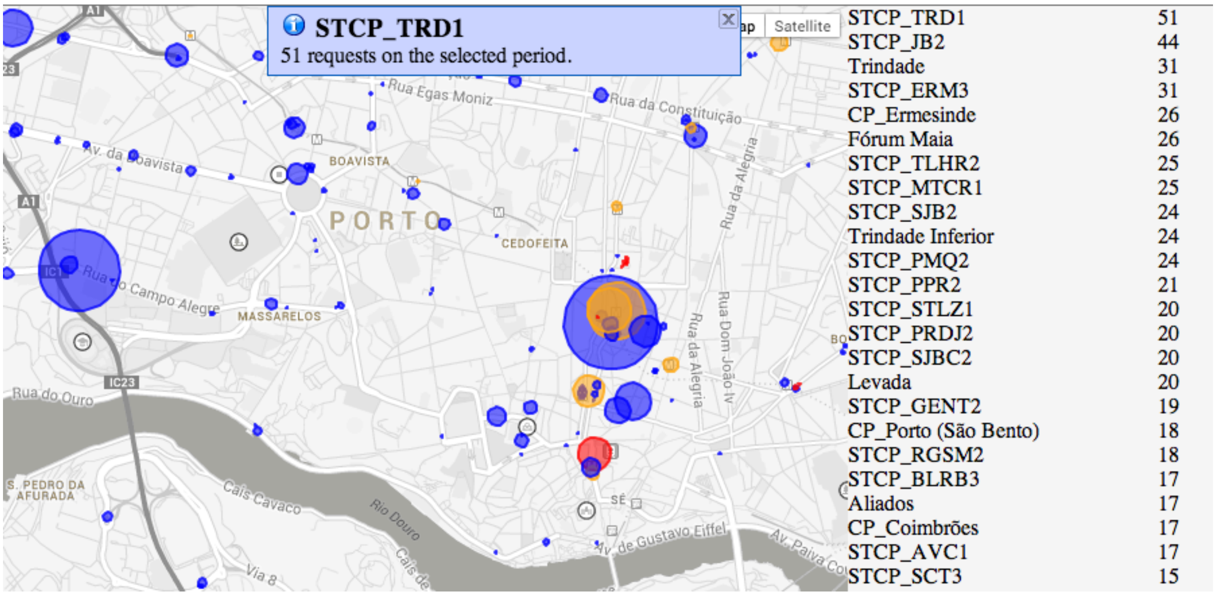


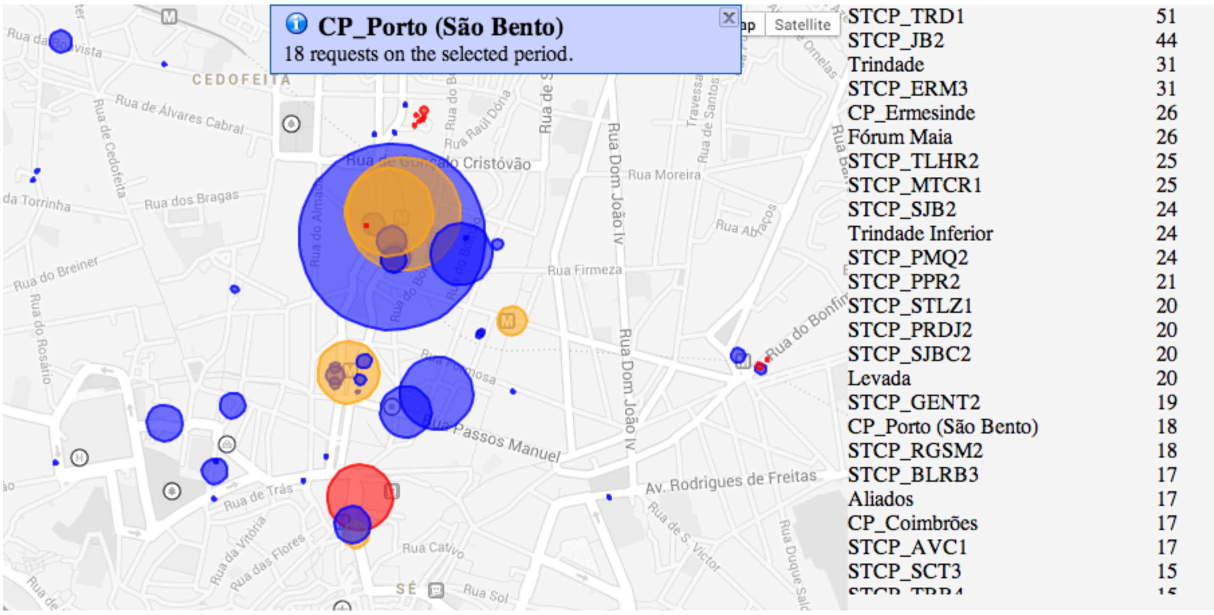
Heat maps


Santa Cruz do Bispo
Leça do Balio
Gueifães
Ermesinde
Águas Santas
Corgo
Parada
Gonçalves
Custóias
Recarei
S. Mamede de Infesta
Baguim do Monte
Forno
Matosinhos
Senhora da Hora
Rio Tinto
Vila Nova de Aldoar
Ramalde
Paranhos
Fânzeres
Manariz
Belavista
Porto
Lordelo do Ouro
Campanhã
Vilar
Foz do Douro
Rio Douro
S. Pedro da Afurada
Lagoa
Gondomar
Rua Bélgica
Valbom
Oliveira do Douro
Vila Nova de Gaia
Canidelo
Madalena
Santo Ovídio
Avintes
Jovim
Av. Liberdade
A28
E1
N208
E82
IP4
A4
IC1
N14
N105
N15
IC23
A3
A20
N12
A1
N108
A43
N209
IC29
A44
IP1
N-222


Rua da Constituição
Rua Egas Moniz
Rua da Constituição
BOAVISTA
PORTO
CEDOFEITA
Rua da Alegria
MASSARELOS
Rua Dom João IV
Rua da Alegria
SÉ
Cais Cavaco
Rio Douro
Av. de Gustavo Eiffel
Via 8

Sized circles

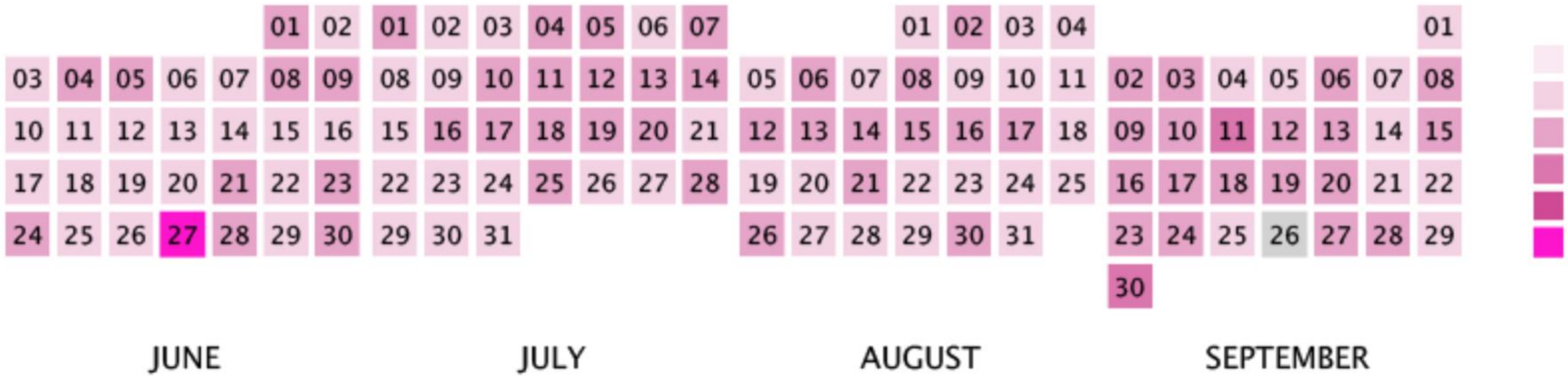




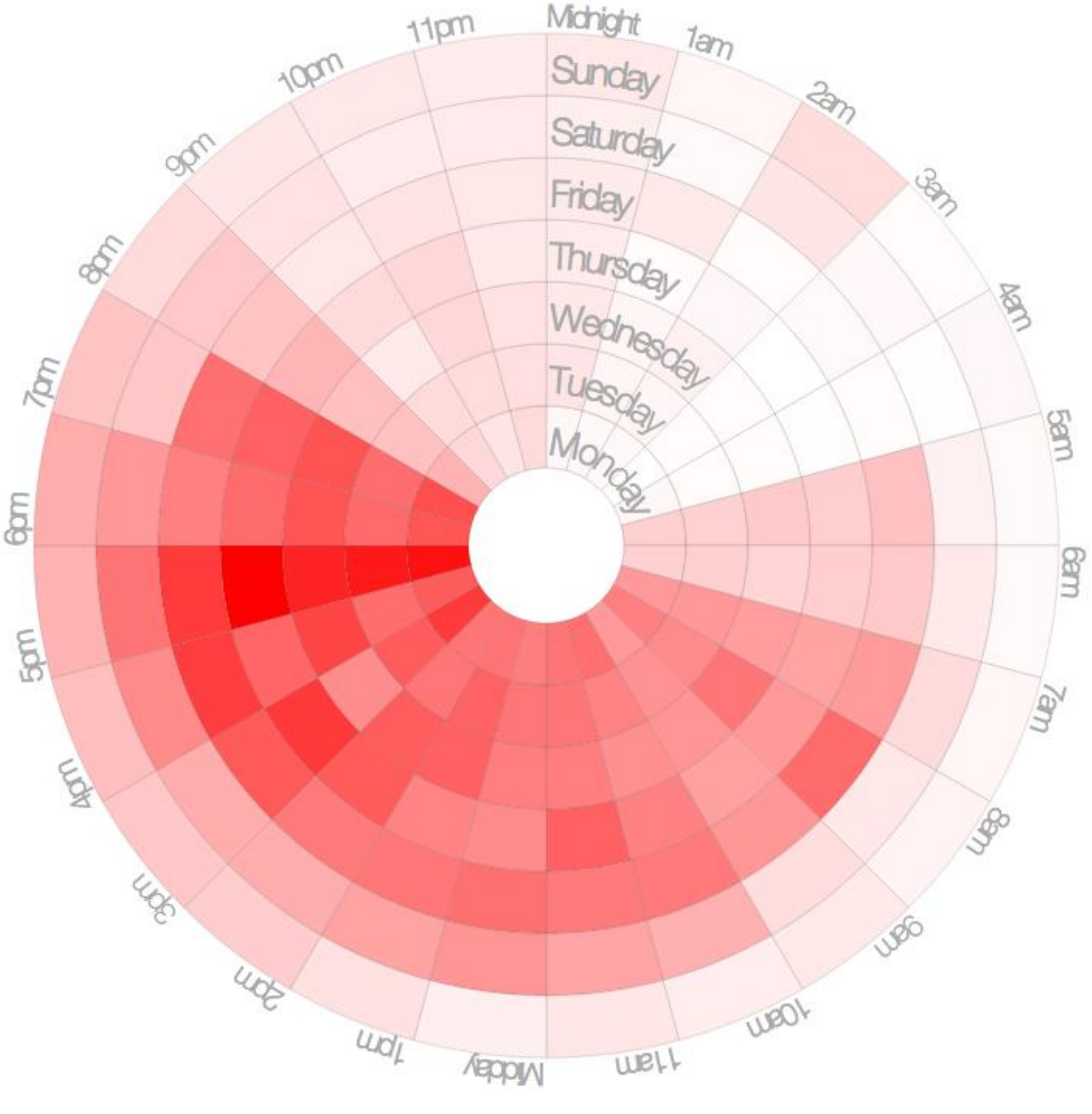
Calendar heat chart

Overview

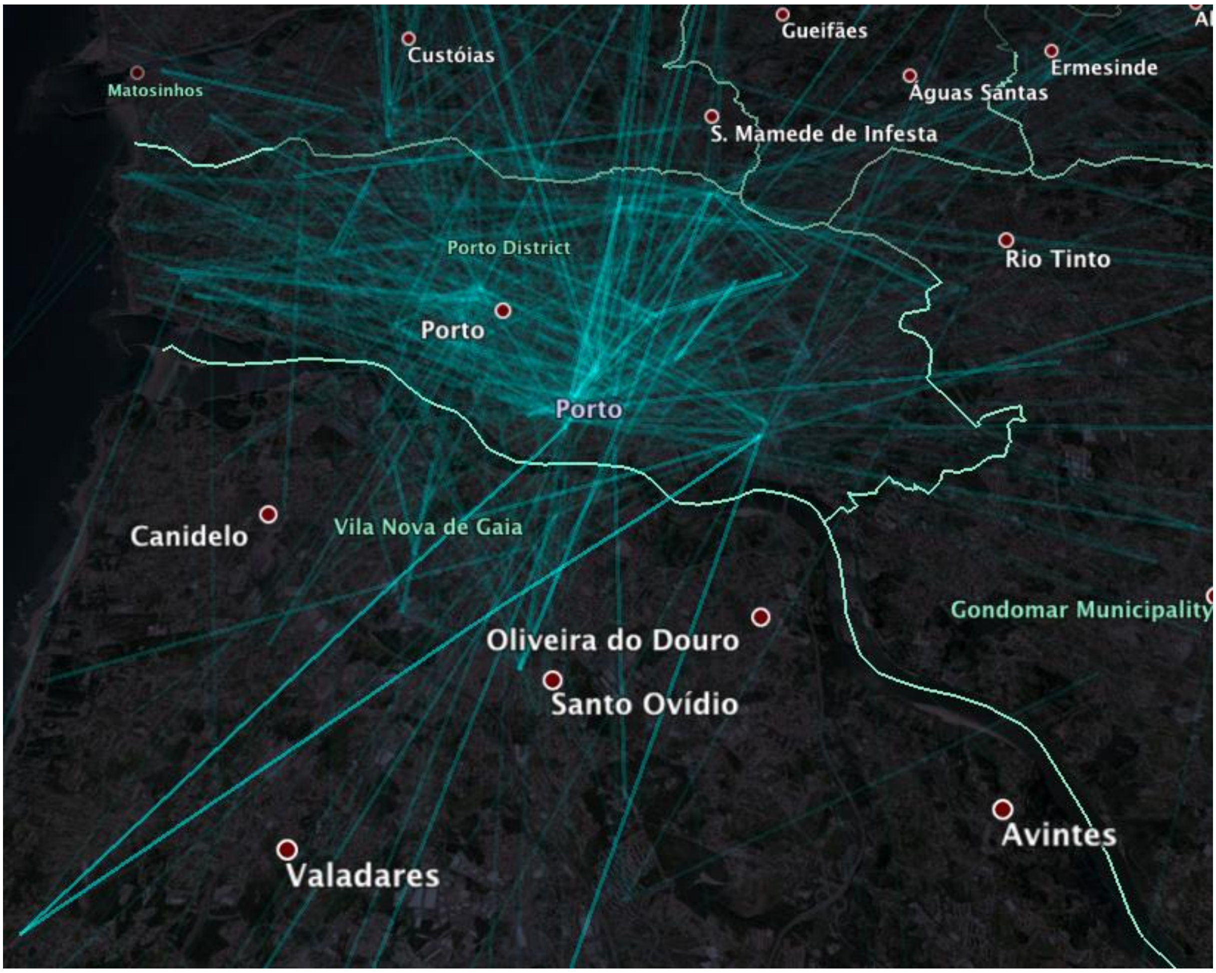
Focus on a specific set of months



Radial heat chart



3D Lines

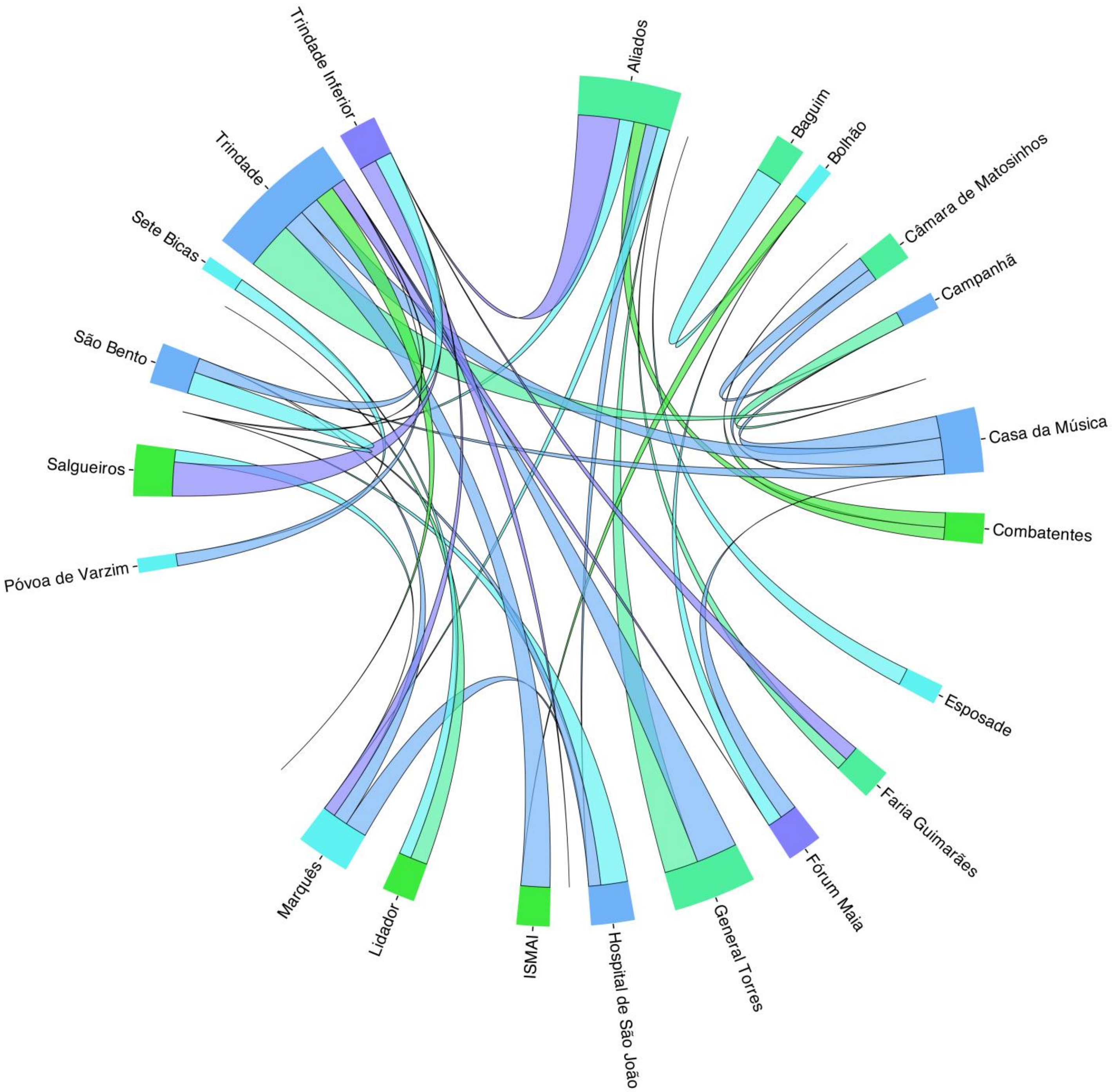


Custóias
Gueifães
Matosinhos
Ermesinde
Águas Santas
S. Mamede de Infesta
Porto District
Rio Tinto
Porto
Porto
Canidelo
Vila Nova de Gaia
Gondomar Municipality
Oliveira do Douro
Santo Ovídio
Avintes
Valadares

Alfena
Póvoa de Varzim
Valongo Municipality
Vila do Conde
Maia
Maia
Rio Tinto
S. Mamede de Infesta
Gondomar Municipality
Porto District
Porto
Porto
Avintes
Matosinhos
Matosinhos
Vila Nova de Gaia
Canidelo
Gulpilhares

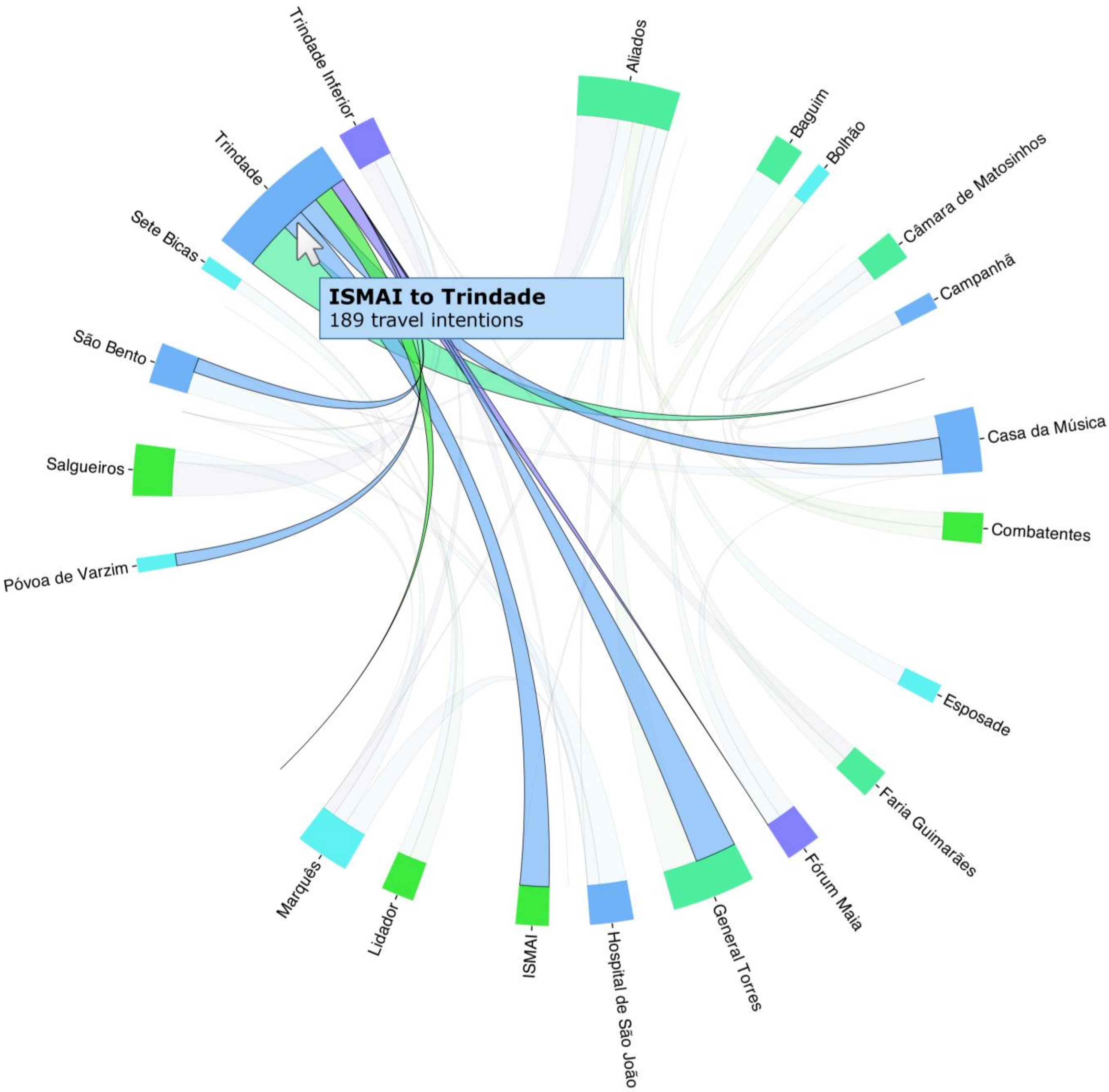
Radial convergence layout

Overview



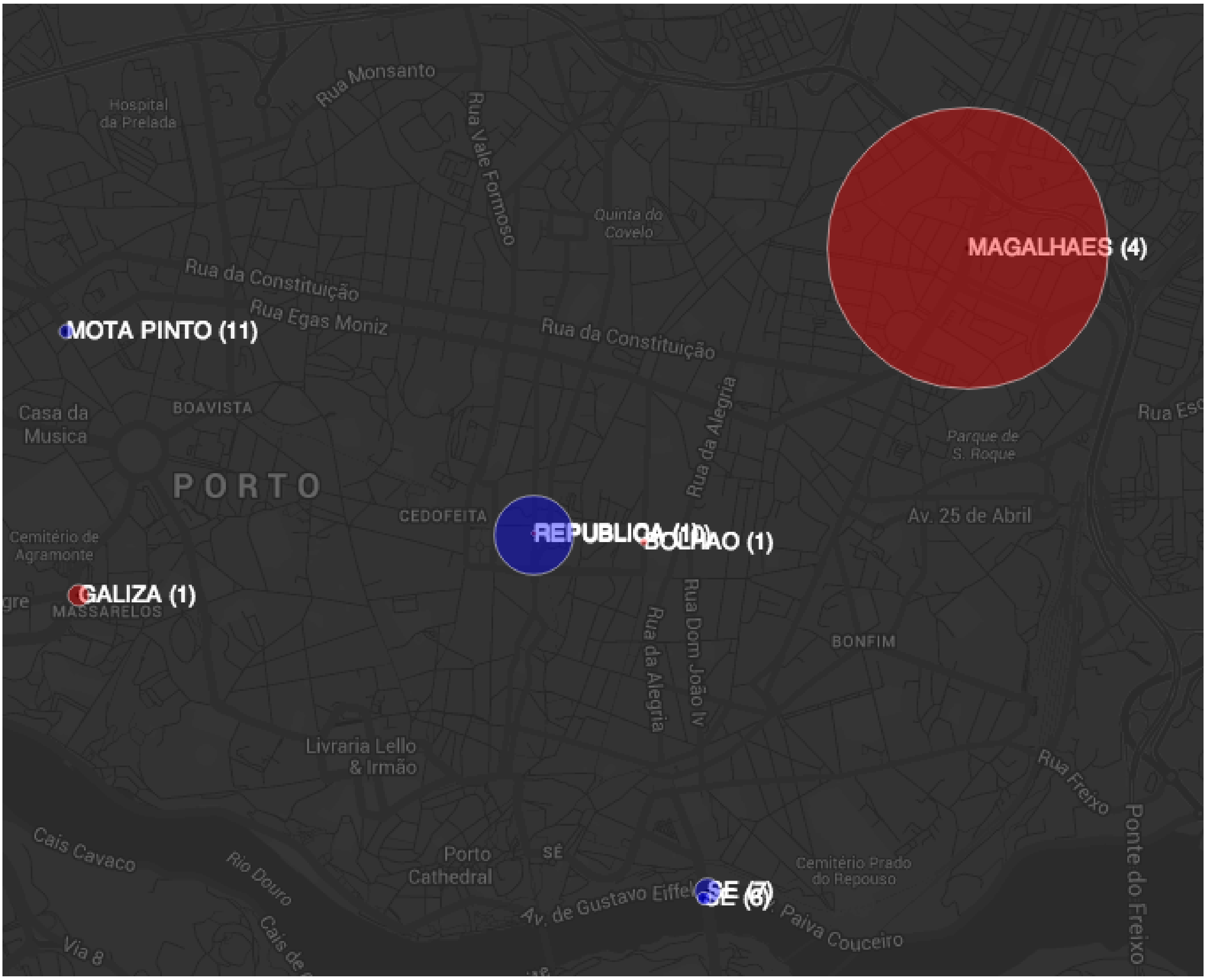
Radial convergence layout

Focus on Trindade stop (Selected chord: ISMAI → Trindade)



Coupled visualisation

Overview



Coupled visualisation

Focus on *Magalhães (4)* sensor